Information Transfer through Quantum Channels

Von der Fakultät für Physik
der Technischen Universität Carolo-Wilhelmina
zu Braunschweig
zur Erlangung des Grades eines
Doktors der Naturwissenschaften
(Dr. rer. nat.)
genehmigte

D i s s e r t a t i o n

von Dennis Kretschmann

aus Braunschweig

1. Referent: Prof. Dr. Reinhard F. Werner
2. Referent: Prof. Dr. Alexander S. Holevo
Eingereicht am: 18. Dez. 2006
Mündliche Prüfung (Disputation) am: 12. März 2007
Druckjahr: 2007

Vorveröffentlichungen der Dissertation

Teilergebnisse aus dieser Arbeit wurden mit Genehmigung der Fakultät für Physik, vertreten durch den Mentor der Arbeit, in folgenden Beiträgen vorab veröffentlicht:

Publikationen:

1. D.Kretschmann, R. F. Werner: *Tema Con Variazioni: Quantum Channel Capacity*; New Journal of Physics **6** (2004) 26 (quant-ph/0311037)

2. D. Kretschmann, R. F. Werner: *Quantum Channels with Memory*; Phys. Rev. A **72** (2005) 062323 (quant-ph/0502106)

3. D. Kretschmann: *Capacity for Quantum Information*; in Eds. J.-P. Françoise, F. G. Naber, S. T. Tsou: *Encyclopedia of Mathematical Physics;* Elsevier, Amsterdam 2006, p. 424

4. D. Kretschmann, D. Schlingemann, R. F. Werner: *The Information-Disturbance Tradeoff and the Continuity of Stinespring's Representation*; quant-ph/0605009 (Mai 2006), eingereicht bei IEEE Trans. Inf. Th.

5. G. M. D'Ariano, D. Kretschmann, D. Schlingemann, R. F. Werner: *Quantum Bit Commitment Revisited: the Possible and the Impossible*; quant-ph/0605224 (Mai 2006), eingereicht bei Phys. Rev. A

Tagungsbeiträge:

A. D. Kretschmann, R. F. Werner: *Quantum Cellular Automata and the Capacity of Quantum Channels with Memory;* Vortrag; Frühjahrstagung der Deutschen Physikalischen Gesellschaft, München, März 2004

B. D. Kretschmann, R. F. Werner: *Quantum Channels with Memory;* Poster; Konferenz Entanglement, Information, and Noise; Krzyzowa, Polen, Juni 2004

C. D. Kretschmann, R. F. Werner: *Quantum Channels with Memory;* Vortrag; Erato Conference on Quantum Information Science, Tokio, Sept. 2004

D. D. Kretschmann, D. Schlingemann, R. F. Werner: *No Perturbation without Measurement;* Vortrag; Frühjahrstagung der Deutschen Physikalischen Gesellschaft, Berlin, März 2005

E. D. Kretschmann, R. F. Werner: *Forgetful Quantum Channels;* Vortrag; IQING4 Konferenz; Paris, Juli 2005

F. D. Kretschmann: *Quantum Bit Commitment Revisited: the Possible and the Impossible;* Vortrag; Young European Physicists Meeting; Budmerice, Slowakei, Okt. 2005

G. D. Kretschmann: *Quantum Bit Commitment Revisited: the Possible and the Impossible;* Vortrag; Frühjahrstagung der Deutschen Physikalischen Gesellschaft, Frankfurt, März 2006

H. D. Kretschmann: *The Information-Disturbance Tradeoff and the Continuity of Stinespring's Representation;* Vortrag; International Congress on Mathematical Physics (ICMP), Rio de Janeiro, Aug. 2006

Drinnen ist es natürlich dunkel, dazu hast du deine Kopflampe,
und von der letzten Sprosse aus steigst du dann praktisch ins Wasser,
oder in den Schmutz, oder zum Beispiel in Schotter.
Du steigst, wie wir sagen, ins Material.

Gerhard Rosner, Kanalarbeiter [Ros04]

Abstract and Overview

This PhD thesis represents work done between Aug. 2003 and Dec. 2006 in Reinhard F. Werner's quantum information theory group at Technische Universität Braunschweig, and Artur Ekert's Centre for Quantum Computation at the University of Cambridge.

Quantum information science combines ideas from physics, computer science and information theory to investigate how quintessentially quantum mechanical effects such as superposition and entanglement can be employed for the handling and transfer of information. My thesis falls into the field of abstract quantum information theory, which is concerned with the fundamental resources for quantum information processing and their interconversion and tradeoffs. Every such processing of quantum information can be represented as a quantum channel: a completely positive and trace-preserving map between observable algebras associated to physical systems. This work investigates both fundamental properties of quantum channels (mostly in Chs. 3 and 4) and their asymptotic capacities for classical as well as quantum information transfer (in Chs. 5 through 8).

Stinespring's theorem is the basic structure theorem for quantum channels. It implies that every quantum channel can be represented as a unitary evolution on an enlarged system. In Ch. 3 we present a continuity theorem for Stinespring's representation: two quantum channels are similar if and only if it is possible to find unitary implementations that are likewise similar, with dimension-independent norm bounds. The continuity theorem allows to derive a formulation of the information-disturbance tradeoff in terms of quantum channels, and a continuity estimate for the no-broadcasting principle. In Ch. 4 we will then apply the continuity theorem to give a strengthened no-go proof for quantum bit commitment, an important cryptographic primitive. This result also provides a natural characterization of those protocols that fall outside the standard setting of unconditional security, and thus may allow secure bit commitment. We present a new such protocol whose security relies on decoherence in the receiver's lab.

Quantum channel capacities quantify the ultimate physical limits for the faithful transmission of information in the presence of decoherence and noise. Ch. 5 reviews the capacities of quantum channels for the transfer of both (private) classical and quantum information, and investigates several variations in the notion of channel capacity.

Most of the results presented in Ch. 5 are limited to memoryless quantum channels, which are characterized by the assumption that consecutive channel inputs are acted on independently. Memory effects are then investigated in detail in Ch. 6. We advertise a model which is sufficiently general to encompass all *causal* automata: every quantum process in which the outputs up to any given time t do not depend on the inputs at times $t' > t$ can be represented as a concatenated memory channel. We also show that most memory channels are *forgetful*, in the sense that those channels in which the effect of the initializing memory dies out as time increases are both open and dense in the set of memory channels. We then explain how all known coding theorems can be generalized from memoryless channels to forgetful memory channels. We also present examples for non-forgetful channels, and derive generic entropic upper bounds on their capacities for (private) classical and quantum information transfer.

Ch. 7 provides a brief introduction to quantum information spectrum methods as a promising approach to coding theorems for completely general quantum sources and channels — without any structural assumptions such as stationarity, ergodicity, or even causality. We present a data compression theorem for general quantum sources and apply these results to ergodic as well as mixed (non-ergodic) sources.

Finally, in Ch. 8 we investigate the continuity of distillable entanglement — another key notion of the field, which characterizes the optimal asymptotic rate at which maximally entangled states can be generated from many copies of a less entangled state. We derive uniform norm bounds for all states with full support, and we extend some of these results to quantum channel capacities.

We start in Ch. 1 with a bird's eye view of quantum information science, motivate the investigation of quantum channels, and explain the results of this work from a broader perspective. Ch. 2 then provides a brief introduction to the formalism and basic tools of quantum information theory, and contains those technical prerequisites that are common to all of the following chapters. More specialized material (which is nevertheless well-known in much of the community) is relegated to the Appendix. This includes entropic information measures and their basic properties (Appendix A), direct sums and quantum-classical hybrid systems (Appendix B), as well as quasi-local algebras (Appendix C).

Contents

Chapter 1

Introduction

The purpose of this Chapter is to present the subject of the thesis in the larger context of quantum information science, and to motivate the investigation of quantum channels, their properties and capacities.

In Sec. 1.1, we will develop in broad brushstrokes a picture of quantum information science, before turning to quantum channels in Sec. 1.2. It goes without saying that this presentation is far from being comprehensive. A more extensive overview of the field can be found in [Pre99, BEZ00, NC00, ATH+01, Key02, Hay06^b]. In Sec. 1.3 we will then briefly highlight the results described in this thesis, and explain how they fit into the bigger picture. As a guide to the reader, Sec. 1.4 explains the structure of the thesis and the interrelations between its individual chapters.

1.1 Quantum Information Science

Quantum information science is a young and interdisciplinary research field at the intersection of physics, information theory, and computer science. It attracts a lot of attention by proposing fascinating new ways of processing information — some of which outperform all known classical techniques. What is more, it sheds new light on the fundamental questions that lie at the heart of quantum physics. So what is quantum information science, and how can it be put to use?

Information is not purely mathematical. Instead it is always carried by physical systems. In traditional information theory, as founded by Claude E. Shannon [Sha48], these systems are assumed to obey the laws of classical physics alone. Usually one considers d-level systems, which can be in any one of d different states. An electrical switch being either on or off is a simple example of a two-level system, and represents one *bit* of information. (This terminology will be explained in Sec. 2.8.)

Quantum information theory deals with those settings in which the carriers are quantum

particles. A typical carrier of quantum information will now be a two-level quantum system, consisting, for example, of the ground state and the excited state of an atom, or the polarization degrees of freedom of a photon. Due to the so-called *superposition principle* of quantum mechanics, quantum systems can not only exist in any of the d individual levels, but also in arbitrary superposition. In analogy to the classical setting, a quantum mechanical two-level system is said to represent one quantum bit, or *qubit*. Following Reinhard Werner's presentation [Wer01], we may now cast a definition:

> *Quantum information is the kind of information that is carried by quantum systems from the preparation device to the measurement device in a quantum mechanical experiment.*

Is quantum information any different from classical information? Much of the success of Shannon's classical theory stems from the fact that it does not seem to make any reference to a specific carrier, and thus applies to information stored in print, the electromagnetic signals sent through a wire, or the bytes on a compact disk. Abstraction from the physical carrier is possible because information can be easily converted back and forth between all these carriers. But when quantum particles are brought into play, conversion is no longer feasible: the so-called *no-teleportation theorem* implies that we cannot usually convert quantum information into classical information and then back into the original quantum information. This shows that quantum information is a truly new kind of information. The no-teleportation theorem can be seen [Wer01] to be intimately related to the impossibility of copying quantum states, as manifested in Wootters's and Zurek's famous *no-cloning theorem* [WZ82], and also to the impossibility of super-luminal communication.

How can one take advantage of this new kind of information? It is probably too early to give a definite answer. However, there are a number of interesting proposals for devices based on the physics of quantum information that might one day outperform all classical techniques.

One of the most well-known, and at the same time most ambitious of these proposals, is *quantum computation.* The basic idea is, loosely speaking, that a quantum computer can be run on a coherent superposition of all possible classical inputs. This is sometimes called *quantum parallelism*, and may lead to a fabulous speed-up for certain tasks, which renders feasible some problems considered intractable by all classical algorithms. A most impressive example is Peter Shor's factorization algorithm [Sho94, Sho97], providing an exponential speed-up over the best known classical algorithms.

Quantum computing could also turn out to be tremendously helpful in the simulation of quantum systems, and might therefore greatly contribute to the understanding of complex quantum phenomena. Simulating general quantum systems is notoriously difficult for classical computers, since the amount of data needed to describe a quantum system grows exponentially with the size of the system — as opposed to classical

systems, which scale linearly. In general, storing the quantum state of a system with n components requires $\sim c^n$ bits of memory, for some constant c, and the specification of unitary operations on such a system requires a further $\sim c^{2n}$ bits. In contrast, a quantum computer could possibly perform the same simulation with only $\sim n$ qubits, allowing the detailed investigation of quantum mechanical systems that are believed to be intractable on any classical computer. This was pointed out by Feynman as early as 1982 [Fey82], and independently by Manin [Man80, Man99].

Another example impressively demonstrating the potentials of quantum information, in fact a technique much closer to realization in everyday life than quantum computation, is *quantum cryptography*: The basic idea is to exploit the quantum mechanical principle that observation in general disturbs the system being observed, so that an eavesdropper will either leave a trace, or else obtain no useful information at all. From this one may conceive protocols to distribute cryptographic keys between two distant parties without any possibility of a compromise, and security guaranteed by the laws of quantum physics. The first such protocol was invented by Charles Bennett and Gilles Brassard in 1984, and has been dubbed *BB84 protocol* [BB84]. A slightly different scheme was proposed independently by Artur Ekert a few years later [Eke91].

The essential resource behind most of the fascinating applications of quantum information theory is *entanglement*, which Erwin Schrödinger in a 1935 paper called "not *one* but rather *the* characteristic trait of quantum mechanics, the one that enforces its entire departure from classical lines of thought" [Sch35[a]]. Entanglement gives rise to a sort of non-classical correlations between quantum systems. Entangled objects behave as if they were connected with each other — no matter how far apart they are. Interest in entanglement, long regarded as a mere curiosity of quantum mechanics, grew when John Bell predicted [Bel64], and experiments later confirmed [ADR82], that entangled quantum systems show behavior impossible in a classical world, even if one would change the laws of physics to try to emulate the predictions of quantum mechanics within a classical framework of any sort.

That entanglement can be put to use in an information theoretic context can be seen most impressively in a process called *entanglement-assisted teleportation:* while we have argued above that teleportation of quantum states is per se impossible, it may be realized if only sender and receiver in addition share an entangled pair of particles [BBC$^+$93]. Entanglement can also assist in the transfer of classical information: if sender and receiver share entangled pairs of particles, they may transfer two bits of information per qubit sent, instead of only one, in a process called *superdense coding* [BW92].

With the discovery of superdense coding and teleportation, entanglement changed its role from an oddity to a resource, which is used up in the process and can be characterized in quantitative terms. This change of perspective has been imported to quantum physics by contact with classical information theory, which has always been concerned

with the interplay and conversion of asymptotic resources. Both superdense coding and teleportation require entanglement in the form of so-called maximally entangled qubit pairs, or *ebits*. They also work with other entangled states, but then the process becomes less efficient. Entangled states produced in the lab and distributed over long distances typically come with imperfections, but can often be converted into a smaller number of ebits with the help of local quantum operations and classical communication between the sender and the receiver. The maximal asymptotic rate at which noisy states may be upgraded to maximally entangled states is called *distillable entanglement*, and serves as a useful quantitative measure of the amount of entanglement. Other tasks in which entanglement appears suggest different ways of quantifying entanglement, and still other measures have been introduced simply because they appear natural from a mathematical point of view, or provide useful bounds on the more operational measures.

The investigation of the zoo of entanglement measures has blossomed into an extremely rich field and has led to significant progress in the conceptual and quantitative understanding of bipartite entanglement. Entanglement in multipartite systems has a much richer structure and still presents many challenges for research, but nevertheless has already found fruitful applications: the investigation of entanglement properties of infinitely extended lattices such as spin systems of coupled harmonic oscillators has recently led to the development of novel simulation techniques for quantum many-body systems and resulted in useful generalizations of the density matrix renormalization group (DMRG) methods [VPC04]. So it seems that insights from quantum information science are already beginning to make a significant and lasting impact on other fields of physics.

So far we have focused almost entirely on the theory side of quantum information science. But can all these fascinating effects be seen in the lab and exploited for information processing? The experimental realization of quantum circuits, algorithms, and communication devices has proven extremely challenging. However, spectacular progress in the manipulation of individual quantum systems has been made in recent years, and has already resulted in the implementation of several basic quantum information processors. Since the first proof-of-principle demonstration by Bennett *et al.* [BBB$^+$92], quantum cryptography has quickly evolved into a thriving experimental area and even a commercial proposition[1]. Within a decade, it might well be possible to place sources of entangled photons on satellites to establish global quantum communication and cryptography networks.

The next application on the horizon is a quantum simulator, which would only require rather modest control of a few dozen qubits to investigate interesting physical phenomena that no classical computer can simulate. Quantum simulation could prove helpful in the development of new materials, the accurate description of chemical compounds

[1] Visit `www.idquantique.com`, `www.magiqtech.com`, or `www.elsag.it` for more information on commercial quantum cryptography.

and reactions, or even in a better understanding of superconductivity and quantum chromodynamics.

Much more stringent in comparison are the demands for universal scalable quantum computing, which requires the coherent control and manipulation of possibly hundreds of individual qubits, with formidable accuracy. In principle we know how to build such a quantum computer: we start with simple quantum logic gates and connect them up into quantum networks. However, the more interacting qubits are involved, the harder it tends to be to engineer the interaction that would display the desired quantum computation. Several promising model systems for full-scale quantum computation have been identified, including nuclear magnetic resonance [DiV95], trapped ions [CZ95], neutral atoms and cavity QED [Mon02], or quantum dots [LD98], to name just a few. A number of approaches have demonstrated basic sets of gate operations, and for some of those systems quantum computation with a few qubits has meanwhile been realized in the lab. Shor's algorithm has successfully factorized the number $15 = 3 \times 5$ on an NMR quantum computer [VSB$^+$01]. But great challenges remain, and it is probably too early to decide which techniques will ultimately prevail and how a universal quantum computer will look like.

Whenever we handle quantum systems, we will inevitably encounter decoherence effects, which may destroy the precious quantum correlations crucial to all the fascinating techniques described in this Section. Noise and unwanted interactions with the environment become a serious problem for large quantum objects and are precisely what makes a scalable quantum computer so hard to build. As a consequence, error correction is absolutely essential to the field. Of course, noise is a serious problem in the classical setting as well, and there is a well-developed assortment of error-correcting codes to protect classical information against its depredations. However, these techniques are typically based on redundancy and therefore involve the copying of information. But copying of quantum states is forbidden by the no-cloning theorem, so that a direct transfer of these methods to the quantum domain is impossible. The situation looked pretty bleak when clever ideas developed independently by Calderbank and Shor [CS96], and Steane [Ste96] showed how to do quantum error correction without ever learning the states of the quantum system, or needing to clone them. A simple error correction algorithm has been demonstrated experimentally on an NMR quantum computer [CMP$^+$98].

1.2 Quantum Channels

Channels are one of the central notions of both classical and quantum information science. Every processing of quantum information, be it storage or transfer, can be represented as a *quantum channel:* a map $T: \mathcal{A} \to \mathcal{B}$ which transforms states (density matrices) on the sender's end of the channel into states on the receiver's end. As illustrated in Fig. 1.1, we will usually depict such a channel as a box, and the input

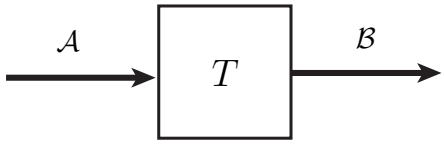

Figure 1.1: *A quantum channel T turns density operators on some (classical or quantum) input system $\mathcal{A}$ into density operators on some output system $\mathcal{B}$.*

and output systems as ingoing and outgoing arrows, respectively. Of course, both the input and the output system may sometimes be composed out of several (classical or quantum) subsystems, and we will frequently consider such channels with more than one input and more than one output system.

The channel concept is completely general and covers every conceivable quantum operation, from state preparations to free or controlled time evolutions to measurements. Examples for physical implementations of quantum channels include optical fibres, through which photons travel from the sender to the receiver, or a sequence of laser pulses addressing a bunch of trapped ions in a quantum computer. However, this is a thesis in abstract quantum information theory, and hence we will usually not be concerned with any particular model system, but rather investigate quantum channels and their properties in full generality. A quantum channel is then simply a map T between observable algebras associated to physical systems — completely characterized by its mathematical properties and somewhat detached from any particular physical implementation.

The abstract approach immediately raises the question which maps T can be interpreted as a quantum channel. The statistical interpretation of quantum mechanics requires that T be linear and positive, so that states are mapped to states. Positivity should still hold true if the channel under consideration is only part of a larger network. In other words, if id_n denotes the noiseless (ideal) channel on an n-dimensional quantum system, then $T \otimes \mathrm{id}_n$ should likewise be positive. The parlance is that T should be *completely positive.* Moreover, the channel T should respect the normalization of density operators, and hence be trace-preserving. In summary, quantum channels are trace-preserving completely positive linear maps $T\colon \mathcal{A} \to \mathcal{B}$ between a sender system $\mathcal{A}$ and a receiver system $\mathcal{B}$. A formal definition along these lines will be presented in Sec. 2.4.

Stinespring's famous dilation theorem [Sti55] is the basic structure theorem for quantum channels: as illustrated later in Fig. 2.1 (cf. Sec. 2.5), it implies that every quantum channel $T\colon \mathcal{A} \to \mathcal{B}$ can be composed from the basic operations of (*i*) tensoring the input with a second system in a specified state (usually called the *ancilla* or *environment* system $\mathcal{E}$), (*ii*) a unitary transformation on the combined input – ancilla system $\mathcal{A} \otimes \mathcal{E}$, followed by (*iii*) a reduction to the output subsystem $\mathcal{B}$. Every quantum channel can

thus be thought of as arising from a unitary evolution on a larger *(dilated)* system. The growth in size due to the additional ancilla system is the price one has to pay for the simplified description of the channel in terms of a unitary rotation. Stinespring's dilation not only provides a very neat characterization of the set of permissible quantum operations, but has also proved a most useful tool in quantum information theory.

Just as a maximally entangled state, a noiseless quantum channel is considered a precious resource in quantum information science. It allows to send quantum information undistorted over some distance, or permits to store it faithfully for some period of time. However, due to detrimental noise effects a perfectly noiseless channel id is seldom available, and the sender and receiver (conventionally called Alice and Bob, respectively) will instead be left with some noisy channel T. Very much in parallel to entanglement distillation protocols described in Sec. 1.1, they may then try to upgrade their noisy resource T with the help of some redundancy and general encoding and decoding operations. The quantum channel capacity $Q(T)$ quantifies how well this can be done: it is the maximal number of qubit transmissions per use of the channel T, taken in the limit of long messages and using error correction techniques asymptotically eliminating all transmission errors. The concept will be investigated in detail in Ch. 5. We can also apply the quantum channel T to send classical information only, resulting in the capacity $C(T) \geq Q(T)$. The investigation of the classical information capacity $C(T)$ has been pioneered by Alexander Holevo [Hol73] and predates the study of quantum information transfer by roughly a quarter century.

1.3 What this Thesis is About: a Preview

This thesis investigates both fundamental properties of quantum channels and their asymptotic capacities for classical and quantum information transfer. It combines techniques from the theory of operator algebras and completely positive maps with ideas from cryptography and information theory. In this Section we will present the results of this work in the larger context of quantum information science.

As explained in Sec. 1.2, Stinespring's dilation theorem is the basic structure theorem for quantum channels. In Ch. 3 we will prove a continuity theorem for Stinespring's representation: two quantum channels T_1 and T_2 are similar iff we can find unitary implementations that are likewise similar. The explicit norm bounds that we derive in Ch. 3 have the welcome property of being completely independent of the dimension of the underlying physical systems. This makes the theorem ideally suited for applications in which this dimension is large or possibly unknown.

The ancilla system in Stinespring's theorem can be interpreted as the environment of the physical system under investigation, and provides a complete description of the external influence on the quantum dynamics. To any channel $T \equiv T_{\mathcal{B}}: \mathcal{A} \to \mathcal{B}$ we may then associate a *complementary* channel $T_{\mathcal{E}}$, in which the roles of the output system

$\mathcal{B}$ and the environment system $\mathcal{E}$ are interchanged. $T_{\mathcal{E}}$ describes the information flow from the input system to the environment. Since complementary channels share a common Stinespring representation, the continuity theorem allows to conclude that two quantum channels are similar iff their complementaries are similar. The complementary channel of the noiseless channel id is the completely depolarizing channel, whose output is completely uncorrelated with the input. Applied to this pair of complementary channels, the continuity theorem allows to conclude that a quantum channel T releases little information to the environment $\mathcal{E}$ iff nearly all the information can be retrieved from the output system $\mathcal{B}$. This result provides quantitative bounds for the information-disturbance tradeoff, which lies at the heart of quantum physics and explains why quantum information behaves so fundamentally different from its classical counterpart. We can also apply the theorem to channels with several outputs, resulting in continuity bounds for the no-cloning principle.

The continuity theorem can also be applied in a cryptographic setting: we show in Ch. 4 how it allows to derive a strengthened no-go proof for classical and quantum bit commitment — a cryptographic primitive involving two mistrustful parties, Alice and Bob. Alice is supposed to submit an encoded bit of information to Bob in such a way that Bob has almost no chance to identify the value of the bit before Alice later decodes it for him, whereas Alice has almost no way of changing the value of the bit once she has submitted it. The setting may appear slightly artificial, but bit commitment does have immediate practical applications and is also known to be an important building block for other cryptographic tasks: it allows to implement secure oblivious transfer, which in turn is enough to establish secure two-party computation. Yet all known bit commitment protocols rely on technological constraints — usually on unproven assumptions that certain computations are hard to perform. Cryptographers have long known that *unconditionally secure* bit commitment (without such technological constraints) cannot be implemented in a classical world.

For quite some time after the invention of unconditionally secure quantum key distribution [BB84, Eke91], there was widespread hope in the community that quantum physics might do the same for bit commitment. The future of quantum cryptography looked very bright until Lo and Chau [LC97], and independently Mayers [May97] realized that under very general premises bit commitment remains impossible even if Alice and Bob are allowed to perform arbitrary quantum operations in their respective labs. However, the Lo-Chau-Mayers no-go result has continually been being challenged. Some researchers have argued that their proof is not general enough to cover all realistic bit commitment scenarios, and several protocols have subsequently been proposed and claimed to circumvent their no-go theorem. These protocols seek to strengthen Bob's position with the help of 'secret parameters' or 'anonymous states', so that Alice lacks some information to cheat successfully.

The continuity theorem allows to derive a strengthened no-go theorem which shows the

insecurity of all such anonymous state protocols. Quantum bit commitment remains impossible for ultimately much of the same reason why purely classical protocols are insecure: if Bob has little information about the value of the committed bit, then he has little means to make sure that Alice sticks to her choice. However, turning this simple idea into a rigorous no-go proof in a quantum setting that allows arbitrary rounds of local operations and classical and quantum communications is far from straightforward. The no-go proof that we present in Ch. 4 also provides a natural classification of those protocols that fall outside the standard setting of unconditional security, and hence may support secure bit commitment. We present a new such protocol whose security relies on decoherence in Bob's lab. Hence, paradoxically, hampering the weaker partner can sometimes lead to successful protocols. Our scheme explores an interesting quantum mechanical effect that has no classical analogue: the distinction between the local erasure of information and the destruction of quantum correlations. Once again, entanglement proves a precious resource.

Ch. 5 investigates capacities for classical and quantum information transfer in a mixture of review and new results. As explained in Sec. 1.2, channel capacities are one of the central notions of the field and quantify the ultimate physical limits for the amount of information that can be sent undistorted through noisy quantum channels. We present several variations in the notion of quantum capacity, and show that they lead to equivalent definitions.

In remarkable contrast to Shannon's purely classical theory, quantum channel capacities (for both classical and quantum information) can be significantly enhanced by additional resources, such as classical side communication and free entanglement shared between sender and receiver. Quantum information theory (sometimes also called *quantum Shannon theory*, since the former term is frequently applied to the theoretical branch of quantum information science as a whole) is concerned with the interplay and tradeoff of all these basics resources, which are in one-to-one correspondence with elementary information-theoretic tasks: sending a unit of quantum (qubit) or classical (cbit) information, sharing the random outcome of a fair coin toss (rbit), or sharing a maximally entangled state (ebit) are the most important ones. Quantum key distribution shows how ebits allow to establish private correlations between two distant parties, but we can also consider private resources directly, such as sending a private classical bit (pbit) or sharing a bit of secret random key (kbit). All these resources are known to be incomparable, but their rates of interconversion are only partly understood. Due to the abundance of inequivalent resources, quantum information theory is much richer than its classical counterpart — a fascinating field with plenty of room for stunning discoveries [DHW05].

Like most of the work in quantum information theory, the overview we present in Ch. 5 is restricted to *memoryless* channels, in which consecutive channel inputs are acted on independently. Mathematically, this means that messages of n symbols are processed

by the tensor product channel $T^{\otimes n}$. The assumption of having uncorrelated noise considerably simplifies the theoretical analysis, but can often hardly be justified. In Ch. 6 we present a general model for quantum channels with memory. We then prove a structure theorem which shows that despite its simplicity, our model is sufficiently general to encompass every reasonable physical process: every *causal automaton*, which turns an infinite string of input states into an infinite string of output states in such a way that outputs up to some time t do not depend on inputs at times $t' > t$ can be represented as a concatenated memory channel. Capacities for memory channels can be defined along the lines familiar from the memoryless setting described in Ch. 5. However, unlike in the memoryless case we also need to specify how to handle the initial and final memory state of each block. In particular, we need to distinguish setups in which either the sender or a malicious third have control of the initial memory state. In general, these distinctions lead to inequivalent capacity concepts. Luckily, these additional complications do not usually occur: we will show that for most memory channels the influence of the initializing memory dies out as time increases. The set of such *forgetful* channels is open and dense in the set of memory channels, so the non-forgetful channels are the exceptional cases. We also show in Ch. 6 how coding theorems can be transferred from the memoryless setting to derive entropic expressions for the (private) classical and quantum capacity of forgetful quantum channels. For non-forgetful channels, only entropic upper bounds on the channel capacities are known.

The results described in Ch. 6 are limited to stationary causal channels and rely to a large extend on techniques originally developed for memoryless channels. A completely different approach to coding for non-i.i.d. states and channels is presented in Ch. 7: the so-called *quantum information spectrum* methods do not require any structural assumptions such as stationarity, ergodicity, or even causality, and in principle allow to derive coding theorems for completely general quantum sources and channels. The necessary techniques are under active development and still hold many challenges for future research — even in the purely classical case. The main result in Ch. 7 is a quantum data compression theorem, which gives the ultimate bounds for the faithful compression and decompression of quantum information.

Finally, in Ch. 8 we come back to entanglement as the basic resource in quantum information science. As explained in Sec. 1.1, distillable entanglement quantifies the maximum asymptotic rate at which ebits may be distilled from many copies of a noisy quantum state. In Ch. 8 we show that distillable entanglement is uniformly continuous on the set of states with full support — and probably on the entire state space. Hence, if two states are almost indistinguishable then they contain similar amounts of distillable entanglement. Since any state preparation in the lab necessarily involves some finite errors, such a continuity property is crucial for an unambiguous operational interpretation of the distillable entanglement. Our results also imply that the quantum channel capacity assisted by two-way classical side communication is continuous in the neighborhood of all channels with vanishing capacity.

1.4 A Guide to the Reader

It will be evident from the preview given in Sec. 1.3 that the following chapters contain results from rather disparate subfields of quantum information science: the theory of completely positive maps, quantum cryptography, quantum Shannon theory, and entanglement theory — all combined under the overarching concept of quantum channels and their properties. Most chapters concentrate on one single topic and are rather self-contained. Ch. 4 on quantum bit commitment clearly depends on the results of Ch. 3 on the continuity of Stinespring's representation. The final three chapters all rely to a lesser extent on the general theory of asymptotic resources, as introduced in Ch. 5.

The general theory of states and channels that forms the common framework for all these results is explained in some detail in Ch. 2. More specialized material that is only relevant to individual sections can be found in the Appendix. Appendix A gives an overview of the most important entropic information measures that appear as asymptotic rate functions in Chs. 5 through 8. Appendix B contains a brief summary of direct sums and quantum-classical hybrid systems, tailored towards the general description of quantum bit commitment protocols in Ch. 4. Finally, Appendix C collects the necessary background on the description of quantum spin chains in terms of quasi-local algebras and is essential to the understanding of the structure theorem for quantum memory channels in Ch. 6. The chart in Fig. 1.2 illustrates the structure of the thesis and the interrelations between its individual chapters.

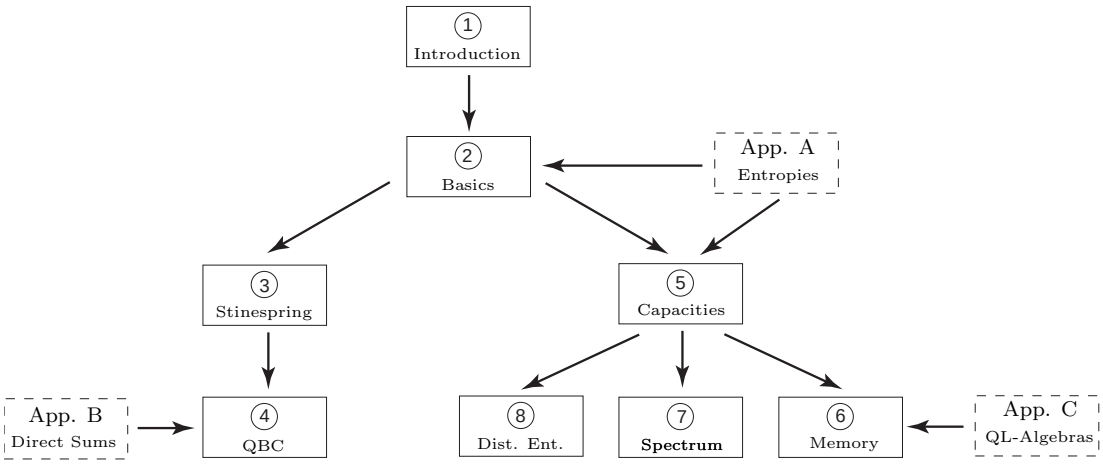

Figure 1.2: *How to read this thesis — the individual chapters and their interdependence.*

Chapter 2

Basic Concepts

In this Chapter we will briefly review the mathematical framework in which the theory of quantum information (and, in fact, classical information as well) are formulated. Though later we will be primarily concerned with quantum information, our description applies to both the classical and the quantum setting. As a consequence, on the one hand the theory comprises *state preparations* and *measurements* in a unified way, and on the other hand, also applies to the transfer of classical information via quantum channels as well as via purely classical channels.

We will start in Sec. 2.1 by explaining how to describe *physical systems* serving as carriers of information, and the preparations and measurements performed on these systems. Composite systems will be investigated in Sec. 2.2, and their classical and quantum correlations will be studied in Sec. 2.3. *Channels* will then be rigorously introduced in Sec. 2.4 as a comprehensive concept of information transfer and processing. Sec. 2.5 is entirely devoted to Stinespring's dilation theorem. When applied to quantum states, Stinespring's theorem yields the GNS-representation, which is instrumental in the proof of the fundamental structure theorem for observable algebras. This will be explained in Sec. 2.6. Finally, in Sec. 2.7, we will show how to appropriately evaluate the distance between quantum channels, as well as between quantum states. In fact, the term *information* itself requires some explanation, which will be given in Sec. 2.8.

We will restrict the discussion to those fundamental concepts that provide the common basis for all the later chapters. More specialized material that will only be instrumental to certain sections will be relegated to later sections and the appendix. This includes an overview of entropic information measures (Appendix A), direct sums and their role for the characterization of algebraically encoded classical information (Appendix B), as well as the description of quantum spin chains in terms of quasi-local algebras (Appendix C).

Our presentation in this Chapter will neither be exhaustive nor completely self-contained. For the sake of brevity, proofs are generally omitted, and we usually restrict the discussion to finite-dimensional quantum systems, since this is the standard setting for

most of quantum information science. A more complete overview of the mathematical foundations of quantum information theory may be found in [Wer01] and [Key02], as well as in the textbooks by Nielsen and Chuang [NC00] and Hayashi [Hay06^b].

2.1 Systems

Classical as well as quantum information theory are statistical (as opposed to deterministic) theories, and their predictions can only be tested if the same experiment is repeated again and again, and the relative frequencies of the outcomes are calculated. In these experiments, physical systems serve as the carriers of information. These systems obey the laws of either classical or quantum physics, and the information theory that results will accordingly be either classical or quantum. Any statistical experiment consists of two different types of procedures: first, a *preparation* procedure, in which a physical system is prepared in a certain *state*, and second, a *registration* procedure, in which a certain *observable* is measured.

A mathematical description of such a twofold setup then naturally consists of two sets, $\mathcal{S}$ and $\mathcal{E}$, and a map

$$\mathcal{S} \times \mathcal{E} \ni (\varrho, A) \mapsto \varrho(A) \in [0,1]. \tag{2.1}$$

The elements of $\mathcal{S}$ describe the states, i. e., preparations of the system, while the elements of $\mathcal{E}$ are the so-called *effects*, which represent all yes/no measurements that can be performed on the systems. The probability to obtain the result "yes" when measuring the effect A on a system prepared in the state ϱ is then given by $\varrho(A)$.

2.1.1 Observable Algebras

The systems we want to describe are either classical or quantum, or hybrids composed of a classical part and a quantum part. A mathematical framework to cover all these settings is the theory of operator algebras, as presented, for instance, in Bratteli's and Robinson's two-volume text [BR87, BR97].

Every system can be completely characterized by its observable algebra $\mathcal{A}$, which can be an arbitrary C^*-algebra with identity $\mathbb{1}_{\mathcal{A}}$. The operations making up the abstract structure of C*-algebras are inspired by those known from algebras of bounded operators $\mathcal{B}(\mathcal{H})$ on some Hilbert space $\mathcal{H}$. In fact, every such operator algebra is a C*-algebra, and conversely, every abstract C*-algebra is isomorphic to a norm-closed self-adjoint algebra of bounded operators on some suitably chosen Hilbert space. More details on this fundamental structure theorem for C*-algebras will be provided in Sec. 2.6 below.

A C*-algebra $\mathcal{A}$ is a vector space on the complex numbers $\mathbb{C}$ which is equipped with a product $\mathcal{A} \times \mathcal{A} \ni A \times B \mapsto A\,B \in \mathcal{A}$. The product is assumed to be distributive and associative, but not necessarily commutative. In addition, $\mathcal{A}$ has an *adjoint* operation

(also called *star operation* or *involution*) $\mathcal{A} \ni A \mapsto A^* \in \mathcal{A}$. This is *conjugate linear* (or *anti-linear*), i. e., $(\alpha A + \beta B)^* = \overline{\alpha} A^* + \overline{\beta} B^*$ for all $A, B \in \mathcal{A}$ and $\alpha, \beta \in \mathbb{C}$, and has the properties $A^{**} = A$ and $(AB)^* = B^* A^*$. Physicists often write A^+ or $A^\dagger$ instead of A^*.

Besides, there is a *norm* $\| \cdot \|_\infty$ on $\mathcal{A}$ which associates a non-negative number $\|A\|_\infty$ to every $A \in \mathcal{A}$ such that $\|A\|_\infty = 0$ implies $A = 0$. With respect to the algebraic properties of $\mathcal{A}$, the norm satisfies the equality $\|\alpha A\|_\infty = |\alpha| \, \|A\|_\infty$, the *triangle inequality* $\|A + B\|_\infty \leq \|A\|_\infty + \|B\|_\infty$, and the *product inequality* $\|AB\|_\infty \leq \|A\|_\infty \, \|B\|_\infty$ for all $A, B \in \mathcal{A}$ and $\alpha \in \mathbb{C}$. In addition, we have $\|A^* A\|_\infty = \|A\|_\infty^2$.

An *identity* $\mathbb{1}_\mathcal{A}$ of a C*-algebra $\mathcal{A}$ is an element of $\mathcal{A}$ such that $\mathbb{1}_\mathcal{A} \, A = A = A \, \mathbb{1}_\mathcal{A}$ for all $A \in \mathcal{A}$. A C*-algebra can have at most one identity. However, not all algebras come equipped with an identity. The absence of an identity can complicate the structural analysis, but these complications are always easily avoided by embedding $\mathcal{A}$ in a larger algebra $\tilde{\mathcal{A}}$ which has an identity. Here we will always assume that $\mathcal{A}$ possesses an identity. Unless the algebra is identically zero, we then have $\|\mathbb{1}_\mathcal{A}\|_\infty = 1$.

The *commutant* $\mathcal{A}'$ of a C*-algebra $\mathcal{A} \subset \mathcal{B}(\mathcal{H})$ is defined as the sub-algebra of all operators $B \in \mathcal{B}(\mathcal{H})$ that commute with $\mathcal{A}$,

$$\mathcal{A}' := \{B \in \mathcal{B}(\mathcal{H}) \;\mid\; A\,B = B\,A \;\; \forall \, A \in \mathcal{A}\}. \tag{2.2}$$

Quantum information theory is primarily concerned with finite-dimensional systems. The structure theorem implies that every finite-dimensional C^*-algebra can be seen as a direct sum of matrix algebras. A quantum d-level system ($d < \infty$) is given by the choice $\mathcal{A} = \mathcal{M}_d$, which denotes the algebra of complex $(d \times d)$ matrices. Here we have only one direct summand. The famous qubit systems correspond to the choice $d = 2$.

At the other extreme we have the classical systems, for which every summand is one-dimensional, and all operators in $\mathcal{A}$ commute, $\mathcal{A} \subset \mathcal{A}'$. Such an algebra is called *commutative* or *Abelian*. In this case $\mathcal{A}$ can be viewed as the algebra of complex-valued functions on the set $\{1, ..., d\}$. This algebra will be denoted by $\mathcal{C}_d$. A classical bit corresponds to the choice $d = 2$. Sometimes it is notationally convenient to use labels other than $1, ..., d$ for the classical state space. We will then write $\mathcal{C}_X$ for the set of complex-valued functions on some finite set X.

2.1.2 States and Effects

Once the observable algebra is known, there is a systematic way to derive the states, $\mathcal{S}$, the effects, $\mathcal{E}$, and the map $(\varrho, A) \mapsto \varrho(A) \in [0, 1]$.

Recall that an element A of the C^*-algebra $\mathcal{A}$ is called *positive* iff it can be written in the form $A = B^* B$ for some $B \in \mathcal{A}$. We then write $A \geq 0$. By $\mathcal{A}^*$ we conventionally denote the *dual space* of $\mathcal{A}$, i. e., the set of all linear functionals on $\mathcal{A}$. The states are

now given by the positive normalized linear functionals on $\mathcal{A}$,

$$\mathcal{S}(\mathcal{A}) := \{\varrho \in \mathcal{A}^* \mid \varrho \geq 0,\, \varrho(\mathbb{1}_{\mathcal{A}}) = 1\}\,, \tag{2.3}$$

where $\varrho \geq 0$ means that $\varrho(A) \geq 0$ for all $0 \leq A \in \mathcal{A}$. The effects, in contrast, are given by

$$\mathcal{E}(\mathcal{A}) := \{A \in \mathcal{A} \mid 0 \leq A \leq \mathbb{1}_{\mathcal{A}}\}\,, \tag{2.4}$$

and the probability to measure the effect A in the state ϱ is simply $\varrho(A)$.

The sets $\mathcal{S}(\mathcal{A})$ and $\mathcal{E}(\mathcal{A})$ are convex, i. e., if $\varrho, \sigma \in \mathcal{S}(\mathcal{A})$, then $\lambda\,\varrho + (1-\lambda)\,\sigma \in \mathcal{S}(\mathcal{A})$ whenever $\lambda \in [0,1]$, and accordingly for $\mathcal{E}(\mathcal{A})$. A distinguished role is played by the *extremal points* of a convex set, i. e., those points τ that do not admit a proper convex decomposition:

$$\tau = \lambda\,\varrho + (1-\lambda)\,\sigma \Longrightarrow \lambda = 0 \;\vee\; \lambda = 1 \;\vee\; \tau = \varrho = \sigma. \tag{2.5}$$

The extremal points of $\mathcal{S}(\mathcal{A})$ are *pure states*, which contain as little randomness as possible, and those of $\mathcal{E}(\mathcal{A})$ are the so-called *propositions*, which are those effects that register a property with certainty.

2.1.3 Quantum Systems

As noted in Sec. 2.1.1, quantum d-level systems are described by the choice $\mathcal{A} = \mathcal{B}(\mathbb{C}^d) = \mathcal{M}_d$, where by $\mathcal{B}(\mathcal{H})$ we conventionally denote the set of bounded linear operators on the Hilbert space $\mathcal{H}$. The algebra $\mathcal{B}(\mathbb{C}^d)$ is itself a Hilbert space when equipped with the Hilbert-Schmidt inner product, $\langle A|B\rangle := \mathrm{tr}(A^*\,B)$ for all operators $A, B \in \mathcal{B}(\mathbb{C}^d)$. Each linear functional $\varrho \in \mathcal{B}^*(\mathbb{C}^d)$ can be expressed in terms of a trace class operator $\tilde{\varrho}$ such that

$$\varrho(A) = \mathrm{tr}(\tilde{\varrho}\,A) \qquad \forall\; A \in \mathcal{B}(\mathbb{C}^d). \tag{2.6}$$

It is obvious that by Eq. (2.6) each $\tilde{\varrho}$ defines a unique linear functional ϱ on $\mathcal{B}(\mathbb{C}^d)$. On the other hand, starting with a given functional ϱ we can recover the matrix elements of $\tilde{\varrho}$ by $\tilde{\varrho}_{jk} = \mathrm{tr}(\tilde{\varrho}\,|j\rangle\langle k|) = \varrho(|j\rangle\langle k|)$, where $\{|j\rangle\langle k|\}_{j,k=1}^{d}$ denotes the canonical basis of $\mathcal{B}(\mathbb{C}^d)$. A state ϱ is pure iff the corresponding trace-class operator $\tilde{\varrho}$ is a rank one projection: $\tilde{\varrho} = |\psi\rangle\langle\psi|$ for some $\psi \in \mathbb{C}^d$. Positivity of the functional ϱ corresponds to positivity of the operator $\tilde{\varrho}$, and the normalization of ϱ translates into the condition $1 = \mathrm{tr}(\tilde{\varrho}) = \varrho(\mathbb{1}_{\mathbb{C}^d})$.

In the following, for simplicity we will usually drop the tilde, and thus identify the trace class operator $\tilde{\varrho}$ with the corresponding linear functional ϱ. The state space $\mathcal{S}(\mathcal{M}_d)$ can then be readily identified with the set of $(d \times d)$ density matrices, as customary in quantum mechanics.

The neat and useful one-to-one correspondence between states and density operators fails to hold for infinite-dimensional Hilbert spaces $\mathcal{H}$: there are always positive linear

functionals on $\mathcal{B}(\mathcal{H})$ which cannot be represented as trace-class operators [Seg47]. The dual space of the trace-class operators are the compact operators on $\mathcal{H}$, not the full algebra $\mathcal{H}$ (cf. [RS80], Th. VI.26). The states that do allow a tracial representation are called *normal*. A positive functional $\omega\colon \mathcal{B}(\mathcal{H}) \to \mathbb{C}$ is normal iff $\lim_{n\to\infty} \omega(A_n) = \omega(A)$ for every sequence $(A_n)_{n\in\mathbb{N}}$ of norm-bounded increasing operators with least upper bound $A \in \mathcal{B}(\mathcal{H})$ (cf. [Dav76], Ch. 1.6). We usually write $\mathcal{B}_*(\mathcal{H})(\subset \mathcal{B}^*(\mathcal{H}))$ for the normal states on the Hilbert space $\mathcal{H}$.

2.1.4 Classical Systems

As mentioned above, the observable algebra of a classical system is the space $\mathcal{C}_X$ of complex-valued functions on the finite set X, $|X| = d < \infty$. In order to interpret this as an operator algebra acting on a finite-dimensional Hilbert space $\mathcal{H}$, we choose a fixed orthonormal basis $\{|x\rangle\}_{x=1}^d$ in $\mathcal{H} := \mathbb{C}^d$, and we identify the function $f \in \mathcal{C}_X$ with the operator $\hat{f} := \sum_{x=1}^d f(x)\,|x\rangle\langle x| \in \mathcal{B}(\mathcal{H})$. Hence $\mathcal{C}_X$ can be thought of as the algebra of diagonal $(d \times d)$ matrices.

From Eq. (2.4) we see that $f \in \mathcal{C}_X$ is an effect iff $0 \leq f(x) \leq 1 \;\forall\; x \in X$. Physically, we can interpret $f(x)$ as the probability that the effect f registers the elementary event $x \in X$. Besides, $p \in \mathcal{E}(\mathcal{C}_X)$ is a proposition iff $p(x) \in \{0, 1\}$ for all $x \in X$.

Since $\mathcal{C}_X$ is finite-dimensional, it is naturally isomorphic to its dual $\mathcal{C}_X^*$: each linear functional $\varrho \in \mathcal{C}_X^*$ is in one-to-one correspondence with the function $x \mapsto \varrho_x := \varrho(|x\rangle\langle x|)$, and writing shorthand $f_x \equiv f(x)$ we have $\varrho(f) = \sum_{x=1}^d f_x\,\varrho_x$. As in the quantum setting we will identify the function $x \mapsto \varrho_x$ with the functional ϱ, and use the same symbol for both.

Positivity of $\varrho \in \mathcal{C}_X^*$ is equivalent to the requirement that $\varrho_x \geq 0 \;\forall\; x \in X$, and the normalization becomes $1 = \varrho(\mathbb{1}_{\mathcal{C}_X}) = \varrho(\sum_{x=1}^d |x\rangle\langle x|) = \sum_{x=1}^d \varrho_x$. Hence, the state $\varrho \in \mathcal{C}_X^*$ corresponds to a discrete probability distribution $\{\varrho_x\}_{x=1}^d$, and ϱ_x is the probability that the elementary event x occurs when the system is in the state ϱ. More generally, $\sum_{x=1}^d f_x\,\varrho_x$ is the probability to measure the effect f when the system is prepared in the state ϱ. The pure states of the system are the *Dirac measures* $\{\delta_x\}_{x\in X}$, with $\delta_x(|y\rangle\langle y|) := \delta_{xy}$, the Dirac delta function.

2.1.5 Observables

Up to now we have discussed only yes/no measurements. We will now show how to build up from them more general observables taking their values in a finite set X. Such an observable E may be thought of as a map $X \ni x \mapsto E_x \in \mathcal{E}(\mathcal{A})$, where $\mathcal{A}$ is the observable algebra of the system under consideration, and E_x is true if x is measured, and false otherwise. If the measurement is performed on systems in the state ϱ, $p_x := \varrho(E_x)$ denotes the probability to obtain the outcome $x \in X$. Hence $\{p_x\}_{x\in X}$

should be a probability distribution on X, which is to say that E should be a so-called *positive operator-valued measure* on X:

Definition 2.1. (Positive Operator-Valued Measure)
For an observable algebra $\mathcal{A}$ and a finite set X, a collection $E = \{E_x\}_{x\in X}$ of effects in $\mathcal{A}$ is called a positive operator-valued measure (POVM) *on X iff $\sum_{x\in X} E_x = \mathbb{1}_{\mathcal{A}}$ holds. Iff all effects E_x are projections, i. e, $E_x^* E_x = E_x \;\forall\; x \in X$, E is called a* projection-valued measure.

In standard quantum mechanics all observables are described by self-adjoint operators on some Hilbert space $\mathcal{H}$. How does this fit into the concept of observables just developed? By the spectral theorem (cf. [RS80], Th. VIII.6), every self-adjoint operator A on the finite-dimensional Hilbert space $\mathcal{H}$ has the form

$$A = \sum_{\lambda\in\sigma(A)} \lambda\, P_\lambda, \tag{2.7}$$

where $\sigma(A)$ denotes the spectrum of A, i. e., the set of eigenvalues λ of A, and P_λ denotes the projection on the corresponding eigenspace. Consequently, there is a projection-valued measure $P = \{P_\lambda\}_{\lambda\in\sigma(A)}$ associated to every self-adjoint operator A, which is conventionally called *spectral measure* of A. It is characterized by the property that the expectation value $\sum_{\lambda\in\sigma(A)} \lambda\, \varrho(P_\lambda)$ of A in the state ϱ is given by $\varrho(A) = \mathrm{tr}(\varrho\, A)$, as familiar from the standard formulation of quantum mechanics. Hence, the traditional way of defining observables in quantum mechanics nicely fits into our scheme, but only covers the projection-valued case. For this reason, general positive operator-valued measures are sometimes called *generalized observables*.

2.2 Composite Systems

While single qubits are certainly interesting, truly fascinating behavior arises when several qubits are brought together. It is here that entanglement enters the stage, which is one of the characteristic traits of quantum mechanics that make quantum physics so different from classical physics, and is the essential resource behind most of the astonishing features of quantum information science. Since entanglement is a multi-system phenomenon, the description of composite quantum systems is essential to the field, and is precisely what this section is about.

For simplicity we will only be concerned with systems consisting of two subsystems, the generalization to multi-partite systems being straightforward. We will follow the scheme introduced in the previous section and talk about systems and subsystems in terms of their observable algebras, in a way which applies to quantum, classical, and hybrid systems alike.

Given two systems with observable algebras $\mathcal{A}$ and $\mathcal{B}$, the observable algebra of the composite system is then simply given by the tensor product $\mathcal{A} \otimes \mathcal{B}$,

$$\mathcal{A} \otimes \mathcal{B} := \text{span}\,\{A \otimes B \mid A \in \mathcal{A},\, B \in \mathcal{B}\}\,. \tag{2.8}$$

The tensor product is a vector space, and can be promoted to a C^*-algebra by defining

$$(A_1 \otimes B_1)\,(A_2 \otimes B_2) := (A_1\,A_2) \otimes (B_1\,B_2) \qquad \text{and} \tag{2.9}$$

$$(A_1 \otimes B_1)^* := A_1^* \otimes B_1^* \tag{2.10}$$

for operators $A_i \in \mathcal{A}$, $B_i \in \mathcal{B}$, $i = 1, 2$. Thus, $\mathbb{1}_{\mathcal{A}\otimes\mathcal{B}} = \mathbb{1}_{\mathcal{A}} \otimes \mathbb{1}_{\mathcal{B}}$.

The physical interpretation of a composite system $\mathcal{A} \otimes \mathcal{B}$ in terms of states, effects, and general observables is straightforward: when $E_\alpha \in \mathcal{A}$ and $E_\beta \in \mathcal{B}$ are effects on the observable algebras $\mathcal{A}$ and $\mathcal{B}$, respectively, $E_\alpha \otimes E_\beta$ is an effect on the product algebra $\mathcal{A} \otimes \mathcal{B}$, and is interpreted as the joint measurement of E_α on the first and E_β on the second subsystem, where the outcome is interpreted as "yes" if both effects have the outcome "yes". In particular, $E_\alpha \otimes \mathbb{1}_{\mathcal{B}}$ corresponds to measuring the effect E_α on the first subsystem while completely ignoring the second. For any given state $\varrho \in \mathcal{A}^* \otimes \mathcal{B}^*$, we may therefore define the restriction $\varrho_{\mathcal{A}}$ of ϱ to the first subsystem by setting $\varrho_{\mathcal{A}}(A) := \varrho(A \otimes \mathbb{1}_{\mathcal{B}})\ \forall\ A \in \mathcal{A}$, and analogously for $\varrho_{\mathcal{B}}$. In the classical case the probability density $\varrho_{\mathcal{A}}$ is obtained by summing (or integrating) out the $\mathcal{B}$ system. In the (finite-dimensional) quantum case this corresponds to *partial tracing* of the density matrices, where the partial trace, $\text{tr}_{\mathcal{B}}\varrho$, is defined on product states by

$$\text{tr}_{\mathcal{B}}(\varrho_{\mathcal{A}} \otimes \varrho_{\mathcal{B}}) := \text{tr}(\varrho_{\mathcal{B}})\,\varrho_{\mathcal{A}}, \tag{2.11}$$

for $\varrho_{\mathcal{A}} \in \mathcal{A}^*$, $\varrho_{\mathcal{B}} \in \mathcal{B}^*$, and linearly extended to $\mathcal{A}^* \otimes \mathcal{B}^*$.

Before we leave this section, we will explore in a little more detail the product algebras that arise from composition in the classical, quantum, and hybrid case. For two classical factors, $\mathcal{C}_X \otimes \mathcal{C}_Y$ with finite sets X and Y, a basis is given by the tensor products of the basis elements, $\{e_x \otimes e_y \mid x = 1, ..., |X|; y = 1, ..., |Y|\}$, and thus $f \in \mathcal{C}_X \otimes \mathcal{C}_Y$ may be given the expansion

$$f = \sum_{x=1}^{|X|} \sum_{y=1}^{|Y|} f_{xy}\, e_x \otimes e_y\,. \tag{2.12}$$

Consequently, f can be identified with a function f_{xy} on the Cartesian product $X \times Y$, implying $\mathcal{C}_X \otimes \mathcal{C}_Y \simeq \mathcal{C}_{X\times Y}$. In other words, states and observables of the composite system $\mathcal{C}_X \otimes \mathcal{C}_Y$ are, in accordance with classical probability theory, given by probability distributions and random variables on the Cartesian product $X \times Y$.

Very similarly, in the purely quantum case we can expand in matrix units to obtain $(A \otimes B)_{\alpha\alpha'\beta\beta'} = A_{\alpha\alpha'}\, B_{\beta\beta'}$, and hence $\mathcal{B}(\mathcal{H}_{\mathcal{A}}) \otimes \mathcal{B}(\mathcal{H}_{\mathcal{B}}) \simeq \mathcal{B}(\mathcal{H}_{\mathcal{A}} \otimes \mathcal{H}_{\mathcal{B}})$.

Now come the hybrid systems consisting of both a classical and a quantum subsystem. They can be approached in two equivalent ways: Suppose we only know that the first

subsystem is classical, without further assumptions on the nature of the second, i. e., we have at hand a system $\mathcal{C}_X \otimes \mathcal{B}$ with some finite set X and some finite observable algebra $\mathcal{B}$. Then every element $W \in \mathcal{C}_X \otimes \mathcal{B}$ can be expanded in the form

$$W = \sum_{x=1}^{|X|} e_x \otimes B_x \tag{2.13}$$

with $B_x \in \mathcal{B} \ \forall \ x = 1, ..., |X|$. Therefore, $\mathcal{C}_X \otimes \mathcal{B}$ can be represented as the set of $\mathcal{B}$-valued functions on X.

Equivalently, assume we only know that $\mathcal{B} = \mathcal{M}_d$ is the algebra of complex $(d \times d)$ matrices, $d < \infty$, and don't have any further knowledge about the finite-dimensional algebra $\mathcal{A}$. We may then expand $W \in \mathcal{A} \otimes \mathcal{M}_d$ in matrix units to find

$$W = \sum_{\beta=1}^{d} \sum_{\beta'=1}^{d} A_{\beta\beta'} \otimes \ e_{\beta\beta'}, \tag{2.14}$$

where $A_{\beta\beta'} \in \mathcal{A} \ \forall \ \beta, \beta' = 1, ..., d$. We can thus identify $\mathcal{A} \otimes \mathcal{M}_d$ with the space of $(d \times d)$ matrices with entries from $\mathcal{A}$. Employing the relation $e_{\alpha\beta} \, e_{\nu\mu} = \delta_{\beta\nu} \, e_{\alpha\mu}$, it can be readily verified that the product in $\mathcal{A} \otimes \mathcal{M}_d$ corresponds to the usual matrix multiplication, with due care given to the order of factors in products if $\mathcal{A}$ happens not to be Abelian. The adjoint is given by $(A^*)_{\alpha\beta} = (A_{\alpha\beta})^*$ for any $A \in \mathcal{A}$.

In summary, our discussion shows that a hybrid algebra $\mathcal{C}_X \otimes \mathcal{M}_d$ with finite set X and $d < \infty$ can be interpreted either as the algebra of $\mathcal{M}_d$-valued functions on X, or as the algebra of $\mathcal{C}_X$-valued $(d \times d)$ matrices.

2.3 Correlations and Entanglement

In this Section we will investigate in some more detail the structure of states on bipartite quantum systems $\mathcal{A} \otimes \mathcal{B}$, and we will briefly discuss the unfamiliar correlations that can sometimes exist between two spatially separated quantum systems and are responsible for most of the fascinating phenomena in quantum information science.

However, for ease of comparison we start our discussion with the hybrid systems, in which $\mathcal{A} = \mathcal{C}_X$ is a classical $|X|$-level system, and $\mathcal{B}$ is some (classical or quantum or hybrid) observable algebra. We know from Sec. 2.1.4 that there exists some orthonormal basis $\{|x\rangle\langle x|\}_{x=1}^{|X|}$ such that every $A \in \mathcal{A}$ can be given the expansion $A = \sum_x a_x \, |x\rangle\langle x|$. Given a quantum state $\varrho \in \mathcal{A}^* \otimes \mathcal{B}^*$, we can now define quantum states $\varrho_x^{\mathcal{A}} \in \mathcal{A}^*$ and $\varrho_x^{\mathcal{B}} \in \mathcal{B}^*$ by setting

$$\varrho_x^{\mathcal{A}}(A) := \langle x|A|x\rangle \qquad \forall \quad A \in \mathcal{A}\,, \tag{2.15}$$

$$\varrho_x^{\mathcal{B}}(B) := \frac{1}{\lambda_x} \, \varrho(|x\rangle\langle x| \otimes B) \qquad \forall \quad B \in \mathcal{B}\,, \tag{2.16}$$

where $\lambda_x := \varrho(|x\rangle\langle x| \otimes \mathbb{1}_{\mathcal{B}})$. Since ϱ is a quantum state, $\{\lambda_x\}_x$ is a classical probability distribution. For any $A \in \mathcal{A}$ and $B \in \mathcal{B}$ we now have

$$\varrho(A \otimes B) = \sum_{x=1}^{|X|} a_x \, \varrho(|x\rangle\langle x| \otimes B) = \sum_{x=1}^{|X|} \lambda_x \, \varrho_x^{\mathcal{A}}(A) \, \varrho_x^{\mathcal{B}}(B) \,, \tag{2.17}$$

and hence every state $\varrho \in \mathcal{A}^* \otimes \mathcal{B}^*$ can be written in the form $\varrho = \sum_x \lambda_x \, \varrho_x^{\mathcal{A}} \otimes \varrho_x^{\mathcal{B}}$. We summarize this important result as

Proposition 2.2. (Classical-Quantum States)
Every state $\varrho \in \mathcal{A}^* \otimes \mathcal{B}^*$ with a classical system $\mathcal{A} = \mathcal{C}_X$ is of the form

$$\varrho = \sum_{x=1}^{|X|} \lambda_x \, \varrho_x^{\mathcal{A}} \otimes \varrho_x^{\mathcal{B}} \,, \tag{2.18}$$

where $\varrho_x^{\mathcal{A}}$ and $\varrho_x^{\mathcal{B}}$ are quantum states on $\mathcal{A}$ and $\mathcal{B}$, respectively, and $\{\lambda_x\}_x$ is a classical probability distribution.

The states of the form Eq. (2.18) can be thought of as being prepared in the following way: conditioned on the outcome of a classical random generator with distribution $\{\lambda_x\}_x$, Alice and Bob prepare the states $\varrho_x^{\mathcal{A}}$ and $\varrho_x^{\mathcal{B}}$ in their respective (possibly distant) labs. Hence, any state of the form Eq. (2.18) can be prepared by local operations (on Alice's and Bob's respective subsystems) and classical communication (LOCC).

Now if both $\mathcal{A}$ and $\mathcal{B}$ are quantum system, the states on the composite system $\mathcal{A}\otimes\mathcal{B}$ can still be correlated in the way just described. But remarkably, it is no longer true that all states on $\mathcal{A} \otimes \mathcal{B}$ are of this form: If $\varrho \equiv |\psi\rangle\langle\psi|$ is pure and has the form Eq. (2.18), it needs to be a product state, i. e., $|\psi\rangle = |\psi^{\mathcal{A}}\rangle \otimes |\psi^{\mathcal{B}}\rangle$. However, it is easily seen that the two-qubit state $|\Omega\rangle := \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$ cannot be decomposed in this way, and the same is true for most other states. The states that can be given a representation of the form Eq. (2.18) are called *separable*; all other states are called *entangled* [Wer89]:

Definition 2.3. (Separability vs. Entanglement)
Let $\mathcal{A}$ and $\mathcal{B}$ be arbitrary observable algebras. A bipartite state $\varrho \in \mathcal{A}^* \otimes \mathcal{B}^*$ is called separable *(or sometimes* classically correlated*) iff there exists a probability distribution $\{\lambda_x\}_x$ and states $\varrho_x^{\mathcal{A}} \in \mathcal{A}^*$ and $\varrho_x^{\mathcal{B}} \in \mathcal{B}^*$ such that*

$$\varrho = \sum_x \lambda_x \, \varrho_x^{\mathcal{A}} \otimes \varrho_x^{\mathcal{B}} \,. \tag{2.19}$$

Otherwise, the state ϱ is called entangled*.*

If there is only one summand in the representation Eq. (2.19), the state ϱ is usually called *uncorrelated*, or a *product state*. We have seen that separable quantum states show classical correlations which can be interpreted as arising from classical random

generators and local operations alone. (This does not mean that the state under consideration has *actually* been prepared in this way, just that this *could* have been the preparation mechanism.) For entangled states, such an interpretation is no longer feasible. The statistics of an entangled state $\varrho \in \mathcal{A}^* \otimes \mathcal{B}^*$ cannot be described by assigning individual properties to the subsystems $\mathcal{A}^*$ and $\mathcal{B}^*$. In a way, the whole is more than the sum of its parts.

Einstein, Podolsky, and Rosen [EPR35] were the first to realize that quantum physics comes with unfamiliar correlations that have no classical interpretation or counterpart — a realization that led them to reject quantum mechanics as a whole, despite its remarkable success. Schrödinger shared their discomfort, and it was in response to the EPR paper that he coined the term *entanglement* (and its German equivalent, *Verschränkung*) for those strange non-classical correlations [Sch35^a, Sch35^b]. For much of the rest of the century, entanglement was mostly seen as an oddity. This view changed completely with the discovery of entanglement-based quantum key distribution [Eke91], dense coding [BW92], and quantum teleportation [BBC$^+$93]. It was suddenly realized that entanglement could be put to use as a powerful resource, enabling fascinating applications beyond anything that could be done with classical systems alone. With the advent of quantum information science, the study of entanglement has blossomed into an extremely active and rich field. Despite considerable progress in the conceptual understanding of entanglement, some of the basic questions still remain unsolved — even in the bipartite case. We refer to the reviews [Bru02, Wer06^a, Wer06^b, PV07] for an overview of the current status and recent developments.

Def. 2.3 merely allows us to decide whether a given state ϱ is entangled or not (at least in principle — to decide whether ϱ can be given the decomposition Eq. (2.19) is actually a highly nontrivial problem, even for small systems). But as with any resource, we would rather like to have a quantitative measure for the amount of entanglement. A large variety of such *entanglement measures* have been suggested. Some of them have a direct physical interpretation, whereas others are purely axiomatic. *Distillable entanglement* is one of the most important operational measures, and will be discussed in more detail in Ch. 8. We again refer to the reviews [Bru02, Wer06^b, PV07] for an overview of the zoo of entanglement measures.

In this Section we will restrict our discussion to *entanglement ordering,* i. e., statements of the form "state ϱ_1 is more entangled than state ϱ_2". We take this to mean that ϱ_2 can be obtained by applying to ϱ_1 some operation which cannot create entanglement. Again, there are different choices for the class of such operations. A natural candidate are the LOCC operations discussed above: we allow **L**ocal **O**perations in Alice's and Bob's respective labs, which can be arbitrary quantum channels in the sense of Ch. 2.4, as well as free use of a noiseless **C**lassical **C**ommunication channel linking both labs. For this class of non-entangling operations, the ordering relations take on a particularly nice form — at least for pure states: Given two bipartite pure states $|\psi\rangle, |\varphi\rangle \in \mathcal{H}_\mathcal{A} \otimes \mathcal{H}_\mathcal{B}$,

we compute the spectral decomposition of the local restrictions $\varrho^\psi := \mathrm{tr}_\mathcal{B}|\psi\rangle\langle\psi| = \sum_i \lambda_i^\psi |i\rangle\langle i|^\psi$ and $\varrho^\varphi := \mathrm{tr}_\mathcal{B}|\varphi\rangle\langle\varphi| = \sum_j \lambda_j^\varphi |j\rangle\langle j|^\varphi$. The state ψ is said to *majorize* the state φ iff for every k the sum of the k largest eigenvalues of ϱ^ψ is larger than the respective sum of eigenvalues of ϱ^φ. Nielsen [Nie99] has shown that this majorization relation completely characterizes the entanglement ordering under LOCC operations:

Proposition 2.4. (Entanglement Ordering)
The pure state $|\varphi\rangle \in \mathcal{H}_\mathcal{A} \otimes \mathcal{H}_\mathcal{B}$ can be converted to the pure state $|\psi\rangle \in \mathcal{H}_\mathcal{A} \otimes \mathcal{H}_\mathcal{B}$ by means of local quantum operations and classical communication (LOCC) iff $|\psi\rangle$ majorizes $|\varphi\rangle$.

By the results of Lo and Popescu [LP01], it is sufficient to consider LOCC protocols with one-way classical communication only — either from Alice to Bob or vice versa. This is always enough if the input state of the LOCC transformation is pure.

If the state ϱ^φ is completely mixed, $\varrho^\varphi = \frac{1}{d}\mathbb{1}_\mathcal{A}$, the corresponding pure state φ is majorized by all other pure states $\psi \in \mathcal{H}_\mathcal{A} \otimes \mathcal{H}_\mathcal{B}$. Hence, any such pure state $|\psi\rangle$ can be prepared from $|\varphi\rangle$ be means of one-way LOCC operations alone, and by classical mixing the same is in fact true for any state $\varrho \in \mathcal{B}_*(\mathcal{H}_\mathcal{A} \otimes \mathcal{H}_\mathcal{B})$. Such a state $|\varphi\rangle$ is called *maximally entangled*. Any maximally entangled state on d-dimensional quantum systems is locally unitarily equivalent to the state

$$|\Omega_d\rangle := \frac{1}{\sqrt{d}} \sum_{j=1}^{d} |jj\rangle\,. \tag{2.20}$$

The maximally entangled qubit state $|\Omega_2\rangle$ is the essential quantum resource for quantum key distribution [Eke91], dense coding [BW92], as well as quantum teleportation [BBC$^+$93], and plays a major role in the theory as the basic unit of entanglement. As explained in Sec. 1.1, it represents one entangled bit, for short *ebit*.

2.4 Channels

As emphasized in Sec. 1.2, channels cover any processing step in information theory, from preparation to free and controlled time evolution to measurements. The purpose of this section is to provide a unified mathematical framework for the description of all these different operations. The basic idea is to classify each channel T according to the way it modifies subsequent measurements: suppose the channel T converts systems with observable algebra $\mathcal{B}$ into systems with observable algebra $\mathcal{A}$. Then by applying first the channel T and then an effect E on the output system $\mathcal{A}$, we have effectively measured an effect on the $\mathcal{B}$-type system, which will then be denoted by $T(E)$. The channel T is hence completely specified by the map $T\colon\ \mathcal{A} \supset \mathcal{E}(\mathcal{A}) \to \mathcal{E}(\mathcal{B}) \subset \mathcal{B}$, and we will in fact identify the channel with this map.

Note that according to this definition channel maps run in the direction opposite to the direction in which the system travels. Alternatively, we can focus on the dynamics of the states and interpret a channel as a map $T_*\colon \mathcal{B}^* \supset \mathcal{S}(\mathcal{B}) \to \mathcal{S}(\mathcal{A}) \subset \mathcal{A}^*$, taking input states ϱ on the $\mathcal{B}$-system into output states $T_*(\varrho)$ on the $\mathcal{A}$-system.

To distinguish between these maps we will say that T describes the channel in the *Heisenberg picture*, whereas T_* is the *Schrödinger picture* representation. In the statistical interpretation these representations should, of course, coincide, i. e., the probabilities $(T_*\varrho)(A)$ to obtain the result "yes" when measuring the effect A in the state $T_*(\varrho)$, and $\varrho(T(A))$ when measuring the effect $T(A)$ in the state ϱ should agree,

$$(T_*\varrho)(A) = \varrho(T(A)) \qquad \forall\ A \in \mathcal{E}(\mathcal{A}),\ \varrho \in \mathcal{S}(\mathcal{B})\,. \tag{2.21}$$

Which maps, T and T_*, may be interpreted as a channel? Since $(T_*\varrho)(A)$ is linear in A, from Eq. (2.21) we immediately see that T is affine, i. e.,

$$T(\lambda_1 A_1 + \lambda_2 A_2) = \lambda_1 T(A_1) + \lambda_2 T(A_2) \tag{2.22}$$

for each convex linear decomposition of effects in $\mathcal{A}$, and can thus be naturally extended to a linear map on $\mathcal{A}$. By Eq. (2.21) this implies that T_* is also linear. Furthermore, in order to be a channel T must map effects to effects, and thus has to be positive, i. e., $T(A) \geq 0$ whenever $\mathcal{A} \ni A \geq 0$. Measuring the trivial effect $\mathbb{1}_{\mathcal{A}}$ corresponds to counting the number of individual experiments, and is equally trivial as a measurement on $\mathcal{B}$. So T has to be *unit-preserving*, or *unital*: $T(\mathbb{1}_{\mathcal{A}}) = \mathbb{1}_{\mathcal{B}}$. By Eq. (2.21) this is equivalent to T_* being likewise a positive operator with the normalization condition $(T_*\varrho)(\mathbb{1}_{\mathcal{A}}) = \varrho(\mathbb{1}_{\mathcal{B}})$. In the density operator representation of the state space, as introduced in Section 2.1.3, this just says that T_* is trace-preserving: $\operatorname{tr} T_*(\varrho) = \operatorname{tr} \varrho$ for all density operators ϱ.

Finally, we would like to have an operation of running two (or more) channels in parallel. So for two channels $T_i\colon \mathcal{A}_i \to \mathcal{B}_i$ $(i = 1, 2)$ we would then have to require that $T_1 \otimes T_2\colon \mathcal{A}_1 \otimes \mathcal{A}_2 \to \mathcal{B}_1 \otimes \mathcal{B}_2$ is also a channel. Since the identity $\mathrm{id}_n \in \mathcal{M}_n$ on n-level quantum systems is one of the channels we would like to describe, we have to have that $T_1 \otimes \mathrm{id}_n$ is positive for all $n \in \mathbb{N}$, i. e., T_1 is *completely positive*. It is natural to ask whether the distinction between positivity and complete positivity introduces anything new: Obviously, complete positivity of $T\colon \mathcal{A} \to \mathcal{B}$ implies positivity. The converse holds if at least one of the algebras $\mathcal{A}$ or $\mathcal{B}$ is Abelian . So positivity and complete positivity coincide when classical systems are involved. In the quantum case, however, there are maps which, while being positive, are not completely positive. The transpose operation is a prominent example (see Paulsen's text [Pau02] for these and other properties of completely positive maps).

For any two completely positive maps T_1 and T_2, the product $T_1 \otimes T_2$ is defined and again completely positive; so requiring tensorability with the "innocent bystander" id_n suffices to make all parallel channels well-defined. Complete positivity can be defined

in the Schrödinger picture as well as in the Heisenberg picture, and it is easily seen that T_* is completely positive iff T is. Summarizing what we have learned, we have the following definition:

Definition 2.5. (Channel)
A channel *converting systems with observable algebra $\mathcal{B}$ into systems with observable algebra $\mathcal{A}$ is a completely positive unital linear operator $T\colon \mathcal{A} \to \mathcal{B}$.*

We will now briefly revisit the special cases introduced in Section 2.1: A channel converting classical information to $\mathcal{A}$-type information is a channel $T\colon \mathcal{A} \to \mathcal{C}_X$ with some finite set X, and can be given the decomposition

$$T(A) \;=\; \sum_{x=1}^{|X|} \varrho_x(A)\; e_x \qquad \forall\; A \in \mathcal{A}\,, \tag{2.23}$$

where each $\varrho_x : A \mapsto \varrho_x(A) := T(A)(x)$ is a positive normalized functional on $\mathcal{A}$, i. e., a state in the sense of Sec. 2.1.2. Hence, a channel of this type describes a parameter-dependent preparation, or *preparator*.

If in addition the output system is likewise classical, i. e., $T\colon \mathcal{C}_X \to \mathcal{C}_X$, the channel T is completely specified by the $(d \times d)$ matrix $(T_{xy})_{x,y=1}^{|X|}$ of transition probabilities, with T_{xy} describing the probability to receive the symbol y when the symbol x was sent: $T_{xy} := T(e_y)(x)$ for all $x, y = 1, ..., |X|$. In terms of the transition matrix $(T_{xy})_{x,y=1}^{|X|}$, Eq. (2.23) can be rewritten as follows:

$$(Tf)(x) \;=\; \sum_{y=1}^{|X|} T_{xy}\, f_y \tag{2.24}$$

for any $f \in \mathcal{C}_X$ and $x = 1, ..., |X|$.

Dually, a measurement is simply a channel $T\colon \mathcal{A} \to \mathcal{B}$ with classical output algebra $\mathcal{A} = \mathcal{C}_X$, for some finite set X. Then T is completely specified by its values $E_x := T(e_x)$ on the basis $\{e_x\}_{x=1}^{|X|}$ of $\mathcal{A}$, via

$$T(f) \;=\; \sum_{x=1}^{|X|} f_x\, E_x\,, \tag{2.25}$$

and any such map T is a channel iff the maps $E_x \in \mathcal{B}$ are positive and satisfy the normalization condition $\sum_{x=1}^{|X|} E_x = \mathbb{1}_{\mathcal{B}}$, or, in other words, $\{E_x\}_{x=1}^{|X|}$ is an *observable* in the sense of Def. 2.1. Making use of this one-to-one correspondence, both the set $\{E_x\}_{x=1}^{|X|}$ and the channel T itself will henceforth be called *observable*, so to arrive at the rather intuitive statement that observables are exactly the channels extracting classical information from the given system.

An observable describes only the statistics of the measured outcomes, without giving any information about the state of the system after the measurement process. For a

more detailed description we have to consider the post-measurement quantum systems as an additional output, and are thus left with the channel $T : \mathcal{C}_X \otimes \mathcal{A} \to \mathcal{B}$, where X is the (finite) set of measurement outcomes, and $\mathcal{A}$ describes the output quantum system. Davies invented the term *instrument* for channels of this type [Dav76]. They are specified by a number of $\chi := |X|$ completely positive, yet not unit preserving maps $T_x\colon \mathcal{A} \to \mathcal{B}$, such that $T_x(A) := T(e_x \otimes A)$ for all $A \in \mathcal{A}$, $x = 1, ..., \chi$, implying

$$T(f \otimes A) \;=\; \sum_{x=1}^{\chi} f_x \, T_x(A) \qquad \forall \;\; f \in \mathcal{C}_X \;\; \forall \;\; A \in \mathcal{A}\,. \tag{2.26}$$

Ignoring the $\mathcal{A}$-output leaves an observable with $E_x := T_x(\mathbb{1}_{\mathcal{A}})$, $x = 1, ..., \chi$. On the other hand, by ignoring the classical output we are left with the channel $\overline{T} := \sum_{x=1}^{\chi} T_x$, which gives the overall average state change.

A special instrument is a so-called *von Neumann measurement*, associated with a family of χ orthogonal projections $\{P_x\}_{x=1}^{\chi}$, satisfying $P_x\, P_y = \delta_{xy}\, P_x$ and $\sum_{x=1}^{\chi} P_x = \mathbb{1}_{\mathcal{A}}$. They can be easily seen to define an instrument $T\colon \mathcal{C}_X \otimes \mathcal{A} \to \mathcal{A}$ via $T_x(A) := P_x\, A\, P_x \;\; \forall \;\; A \in \mathcal{A}$. Since $T_x\, T_y = \delta_{xy}\, T_x$ for all $x, y = 1, ..., \chi$, repeating the measurement an arbitrary number of times in this case will always give the same output. What von Neumann actually proposed [Neu55] was to choose one-dimensional projections $\{P_x\}_{x=1}^{\chi}$. The general case is sometimes called an *incomplete von Neumann measurement*, or *Lüders measurement*.

2.5 Stinespring's Representation Theorem

Stinespring's famous representation theorem [Sti55, Arv69] is the basic structure theorem for completely positive maps. As explained in Sec. 1.2, it not only provides a neat characterization of the set of permissible quantum operations, but is also a most useful tool in quantum information science. We will start our presentation with maps between arbitrary C^*-algebras, but will later concentrate on the finite-dimensional case that quantum information theory is chiefly concerned with.

Theorem 2.6. (Stinespring Dilation Theorem)
Let $\mathcal{A}$ be a C^*-algebra, and let $T\colon \mathcal{A} \to \mathcal{B}(\mathcal{H})$ be a completely positive linear map. We may then find a Hilbert space $\mathcal{K}$ and a bounded operator $V\colon \mathcal{H} \to \mathcal{K}$ such that

$$T(A) \;=\; V^*\, \pi(A)\, V \qquad \forall \;\; A \in \mathcal{A}\,, \tag{2.27}$$

where $\pi\colon \mathcal{A} \to \mathcal{B}(\mathcal{K})$ is a *-representation, i. e., a linear operator that preserves the algebraic structure: $\pi(AB) = \pi(A)\,\pi(B)$ and $\pi(A^*) = \pi(A)^*$ for all $A, B \in \mathcal{A}$. If T is unital then V is an isometry, $V^*V = \mathbb{1}_{\mathcal{B}(\mathcal{H})}$.

A triple $(\mathcal{K}, \pi, V)$ as obtained in Stinespring's Theorem is usually called a *Stinespring representation* for the completely positive map T. If the closed linear span of $\pi(\mathcal{A})V\mathcal{H}$

equals $\mathcal{K}$, the representation is called *minimal*. Minimal Stinespring representations are unique up to unitary equivalence, in the following sense: Assume that the quantum channel T has a minimal Stinespring representation $(\mathcal{K}, \pi, V)$ as in Eq. (2.27) as well as a further (not necessarily minimal) one $(\mathcal{K}_1, \pi_1, V_1)$ such that

$$T(A) \;=\; V_1^*\, \pi_1(A)\, V_1 \qquad \forall\; A \in \mathcal{A} \tag{2.28}$$

with another Stinespring isometry $V_1 \colon \mathcal{H} \to \mathcal{K}_1$. Since the representation Eq. (2.27) is assumed to be minimal, we conclude that $\dim \mathcal{K} \leq \dim \mathcal{K}_1$, and the prescription

$$W(\pi(A)V\psi) \;:=\; \pi_1(A)V_1\psi \tag{2.29}$$

for $A \in \mathcal{A}$ and $\psi \in \mathcal{H}$ yields a well-defined isometry $W \colon \mathcal{K} \to \mathcal{K}_1$. In particular, by choosing $A = \mathbb{1}_{\mathcal{A}}$ in Eq. (2.29) we see that $WV = V_1$. From the definition of W we immediately find the intertwining relation

$$W\pi = \pi_1 W\,. \tag{2.30}$$

If both representations are assumed to be minimal, we have $\mathcal{K} \simeq \mathcal{K}_1$, and W is indeed unitary. We summarize the uniqueness clause in the following

Theorem 2.7. (Uniqueness of Stinespring's Dilation)
If $(\mathcal{K}, \pi, V)$ and $(\mathcal{K}_1, \pi_1, V_1)$ are two Stinespring representations for the quantum channel $T : \mathcal{A} \to \mathcal{B}(\mathcal{H})$, as in Eq. (2.27) above, we may find a partial isometry $W : \mathcal{K} \to \mathcal{K}_1$ such that

$$\begin{aligned} WV &= V_1, && && (2.31)\\ W^*V_1 &= V && \text{and} && (2.32)\\ W\pi(A) &= \pi_1(A)W && && (2.33) \end{aligned}$$

for all $A \in \mathcal{A}$. In particular, the minimal Stinespring dilation is unique up to unitary equivalence.

For channels $T \colon \mathcal{M}_n \to \mathcal{M}_m$ between finite-dimensional observable algebras, Stinespring's representation can be given the simpler form

$$T(A) \;=\; V^*(A \otimes \mathbb{1}_l)V \qquad \forall\; A \in \mathcal{M}_n\,, \tag{2.34}$$

with the Stinespring isometry $V : \mathbb{C}^m \longrightarrow \mathbb{C}^n \otimes \mathbb{C}^l$. We will henceforth write $(\mathbb{C}^l, V)$ to denote such a representation. The minimal representation comes with a bound on the dilation space, $l \leq n \times m$.

By means of the duality relation Eq. (2.21), in the Schrödinger picture this form of Stinespring's theorem gives rise to the so-called *ancilla representation* of the quantum channel T_*,

$$T_*(\varrho) = \mathrm{tr}_{\mathbb{C}^l}\, V\, \varrho\, V^* \qquad \forall\; \varrho \in \mathcal{B}_*(\mathbb{C}^m)\,. \tag{2.35}$$

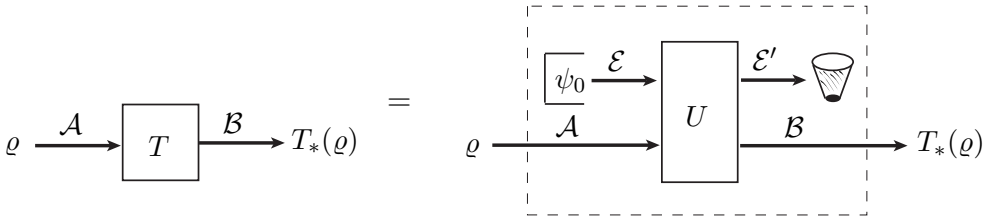

Figure 2.1: *According to Stinespring's representation theorem, the quantum channel T with input system $\mathcal{A}$ and output system $\mathcal{B}$ can be represented in terms of a unitary evolution U on a larger system $\mathcal{A}\otimes\mathcal{E}\simeq\mathcal{B}\otimes\mathcal{E}'$. The environment system $\mathcal{E}$ is prepared in some input state $|\psi_0\rangle$, and the joint input state $\varrho\otimes|\psi_0\rangle\langle\psi_0|$ is then subjected to the unitary dynamics $U(\cdot)\,U^*$. When the environment output $\mathcal{E}'$ is ignored (as symbolized by the waste bin), the resulting output on $\mathcal{B}$ coincides with $T_*(\varrho)$, for all input states ϱ.*

In the physical interpretation of Stinespring's theorem, the dilation space $\mathbb{C}^l$ represents the *environment*. Stinespring's isometry V transforms the input state $\varrho\in\mathcal{B}_*(\mathbb{C}^m)$ into the state $V\varrho V^*$ on $\mathbb{C}^n\otimes\mathbb{C}^l$, which is correlated between the output and the environment. The output state $T_*(\varrho)\in\mathcal{B}_*(\mathbb{C}^n)$ is then obtained by tracing out the degrees of freedom of the environment. Physically, one would expect a unitary operation U instead of an isometric V. However, the initial state of the environment can be considered fixed, effectively reducing U to an isometry, $V\psi:=U(\psi\otimes\psi_0)$ for some fixed initial pure state $|\psi_0\rangle$ of the environment system. Fig. 2.1 illustrates the unitary representation of the quantum channel T.

There is another convenient way of characterizing channels closely related to Stinespring's theorem, the so-called *Kraus form* [Kra83]: by introducing a complete family of one-dimensional projectors $\{|\psi_j\rangle\langle\psi_j|\}_{j=1}^l$ of $\mathbb{C}^l$, and defining the so-called *Kraus operators* $\{t_j\}_{j=1}^l$ in terms of the Stinespring isometry V by $\langle\alpha|t_j|\beta\rangle:=\langle\alpha\otimes\psi_j|V|\beta\rangle$ for all vectors $\alpha\in\mathbb{C}^n$, $\beta\in\mathbb{C}^m$, Eq. (2.34) directly gives the following

Corollary 2.8. (Kraus Form)
Every completely positive linear map $T\colon\mathcal{M}_n\to\mathcal{M}_m$ can be given the form

$$T(A)\;=\;\sum_{j=1}^{N}t_j^*\,A\,t_j\qquad\forall\;\;A\;\in\;\mathcal{M}_n \tag{2.36}$$

with linear operators $t_j\colon\mathbb{C}^m\to\mathbb{C}^n$, and $N\;\leq\;n\,m$.

It is then easily seen from the duality relation Eq. (2.21) that in the Schrödinger representation the channel T has the Kraus form

$$T_*(\varrho)\;=\;\sum_{j=1}^{N}t_j\,\varrho\,t_j^*\qquad\forall\;\;\varrho\;\in\;\mathcal{M}_n^*\,. \tag{2.37}$$

2.6 GNS-Representation of Quantum States

A state $\omega\colon \mathcal{A} \to \mathbb{C}$, as defined in Sec. 2.1.2 above, is a unital and positive linear map. Since the range algebra $\mathbb{C}$ is Abelian, we know that it is even completely positive (cf. [Pau02], Th. 3.9), and we may thus conclude from Stinespring's theorem that ω can be given the representation

$$\omega(A) := \langle\Omega|\pi(A)|\Omega\rangle \qquad \forall\, A \in \mathcal{A}\,, \tag{2.38}$$

where $|\Omega\rangle := V(1)$ is a purification of the quantum state ω. If $\omega \in \mathcal{B}_*(\mathcal{H})$ is a trace-class operator, the representation π can always be assumed to be of the form $\pi(A) = A \otimes \mathbb{1}_{\mathcal{H}'}$ with an ancilla space $\mathcal{H}' \simeq \mathcal{H}$. Eq. (2.38) then shows that we may find a pure state $|\Omega\rangle \in \mathcal{H} \otimes \mathcal{H}'$ such that

$$\omega = \mathrm{tr}_{\mathcal{H}'}|\Omega\rangle\langle\Omega| \tag{2.39}$$

holds by means of the duality relation Eq. (2.21). Eq. (2.38) is usually called the GNS-representation of quantum states, after Gelfand and Naimark [GN43], and Segal [Seg47]. The GNS theorem can be applied to prove the basic structure theorem for C^*-algebras:

Theorem 2.9. (Structure Theorem)
Every C-algebra $\mathcal{A}$ is isomorphic to a norm-closed self-adjoint algebra of bounded operators on a Hilbert space.*

The idea of the proof is to construct for each state ω of $\mathcal{A}$ the corresponding GNS representation $(\mathcal{K}_\omega, \pi_\omega, V_\omega)$, and then to form the so-called *universal representation* by setting

$$\mathcal{K} := \bigoplus_\omega \mathcal{K}_\omega \qquad \text{and} \qquad \pi := \bigoplus_\omega \pi_\omega. \tag{2.40}$$

The existence of sufficiently many states is guaranteed by the Hahn-Banach extension theorem. The details are spelled out in Section 2.3 of [BR87].

2.7 Distance Measures

It will be evident from the preview in Sec. 1.3 that all of the topics we will address in later sections require the comparison of different quantum channels, or different quantum states. There are several natural candidates for such distance measures, which are adapted to different scenarios and reviewed in this Section. We start out with quantum channels.

2.7.1 Distance Measures for Quantum Channels

Assume two channels T_1 and T_2 with common input and output algebras $\mathcal{A}$ and $\mathcal{B}$, respectively. Since these T_i are (in Heisenberg picture) operators between normed

spaces $\mathcal{A}$ and $\mathcal{B}$, the natural choice to quantify their distance is the *operator norm*,

$$\|T_1 - T_2\|_\infty := \sup_{A \neq 0} \frac{\|T_1(A) - T_2(A)\|_\infty}{\|A\|_\infty}. \tag{2.41}$$

The norm distance Eq. (2.41) has a neat operational characterization: it is twice the largest difference between the overall probabilities in two statistical quantum experiments differing only in replacing one use of T_1 with one use of T_2.

However, we also want to allow for more general experiments, in which the two channels are only applied to a subsystem of a larger system. This requires *stabilized* distance measures [GLN05], and naturally leads to the so-called *norm of complete boundedness* (or *cb-norm*, for short) [Pau02]:

$$\|T_1 - T_2\|_{cb} := \sup_{n \in \mathbb{N}} \|\mathrm{id}_n \otimes (T_1 - T_2)\|_\infty \,, \tag{2.42}$$

where again id_n denotes the *ideal* (or *noiseless*) channel on the $(n \times n)$-matrices: $\mathrm{id}_n(M) = M$ for all $M \in \mathcal{M}_n$. Useful properties of the cb-norm include *multiplicativity*, i. e., $\|T_1 \otimes T_2\|_{cb} = \|T_1\|_{cb} \|T_2\|_{cb}$, and *unitality:* $\|T\|_{cb} = 1$ for any channel T.

Obviously, $\|T\|_{cb} \geq \|T\|_\infty$ for every linear map T. If either the input or output space is a classical system, we even have equality: $\|T\|_{cb} = \|T\|_\infty$ (cf. [Pau02], Ch. 3). Fully quantum systems generically show a separation between these two norms. However, in the vicinity of the noiseless channel id the operator norm and the cb-norm may always be estimated in terms of each other with dimension-independent bounds [KW04], and can thus be considered equivalent, even when the dimensions of the underlying Hilbert spaces are unknown and possibly large:

$$\|T - \mathrm{id}\,\|_\infty \leq \|T - \mathrm{id}\,\|_{cb} \leq 8 \, \|T - \mathrm{id}\,\|_\infty^{\frac{1}{4}} . \tag{2.43}$$

Examples which show that this equivalence does *not* hold generally will be provided in Sec. 3.4. Thus, in a quantum world correlations may help to distinguish locally akin quantum channels.

Both operator norm and cb-norm are natural distance measures for quantum channels and have a clear operational interpretation in terms of statistical distinguishability. However, due to the limit over all input observables in Eq. (2.41) — and the additional limit over all bystander systems in case of the cb-norm, Eq. (2.42) — they are often tremendously hard to compute. There is considerably interest in distance measures that are easier to handle and evaluate [GLN05]. A convenient choice is the so-called *channel fidelity*,

$$F_c(T) := \langle \Omega | T \otimes \mathrm{id}\,(|\Omega\rangle\langle\Omega|) | \Omega \rangle \,, \tag{2.44}$$

where $|\Omega\rangle = \frac{1}{\sqrt{d}} \sum_{j=1}^{d} |j\rangle \otimes |j\rangle$ is a maximally entangled state on $\mathbb{C}^d \otimes \mathbb{C}^d$. The channel fidelity $F_c(T)$ is a measure for the quantum channel $T \colon \mathcal{B}(\mathbb{C}^d) \to \mathcal{B}(\mathbb{C}^d)$ to preserve the

entanglement with a bystander system, and is closely related to the *average fidelity* of the channel T,

$$\overline{F}(T) := \int \langle\psi|T(|\psi\rangle\langle\psi|)|\psi\rangle \; d\psi, \tag{2.45}$$

where the integral is over the normalized Haar measure:

Proposition 2.10. *For any quantum channel $T\colon \mathcal{B}(\mathbb{C}^d) \to \mathcal{B}(\mathbb{C}^d)$, we have:*

$$\overline{F}(T) \geq F_c(T) \geq \overline{F}(T) - \frac{1}{d}. \tag{2.46}$$

The proof of Prop. 2.10 is immediate from the relation [HHH99, Nie02]

$$\overline{F}(T) = \frac{d\, F_c(T) + 1}{d+1}. \tag{2.47}$$

Prop. 2.10 shows that average fidelity and channel fidelity become equivalent distance measures in the limit of large dimensions, $d \to \infty$. However, neither can be estimated in terms of the cb-norm difference $\|T - \mathrm{id}\,\|_{cb}$ or the standard operator norm $\|T - \mathrm{id}\,\|_\infty$ with dimension-independent bounds [KW04]. Yet all these measures turn out to lead to equivalent definitions of quantum channel capacity. We will come back to this important point in Sec. 5.2.2.

2.7.2 Distance Measures for Quantum States

We have seen in Sec. 2.6 that states are channels with one-dimensional input space, $\mathcal{B} = \mathbb{C}$. Since this is a classical system, there is no need to distinguish between stabilized and non-stabilized distance measures. The so-called *trace norm* $\|\varrho\|_1 = \mathrm{tr}\sqrt{\varrho^*\varrho}$ is a convenient measure for the distance between two density operators. For any two quantum states $\varrho, \sigma \in \mathcal{B}_*(\mathcal{H})$, the trace norm difference $\|\varrho - \sigma\|_1$ is equivalent to the *fidelity* $f(\varrho,\sigma) := \mathrm{tr}\sqrt{\sqrt{\varrho}\,\sigma\sqrt{\varrho}}$ by means of the following Proposition [NC00]:

Proposition 2.11. (Equivalence)
Let $f(\varrho,\sigma) := \mathrm{tr}\sqrt{\sqrt{\varrho}\,\sigma\sqrt{\varrho}}$ denote the fidelity of two quantum states ϱ, $\sigma \in \mathcal{B}_*(\mathcal{H})$. We then have:

$$1 - f(\varrho,\sigma) \leq \frac{1}{2}\|\varrho - \sigma\|_1 \leq \sqrt{1 - f^2(\varrho,\sigma)}\,. \tag{2.48}$$

The fidelity $f(\varrho,\sigma)$ is symmetric in its inputs and unitarily invariant. It never decreases under quantum operations. If $\varrho = |\varphi\rangle\langle\varphi|$ is pure, we have $f(\varphi,\sigma) = \sqrt{\langle\varphi|\sigma|\varphi\rangle}$. The following Proposition, which appears as Lemma 2 in [SR01], amounts to a kind a triangle-inequality for the fidelity, which will prove useful in our discussion of quantum bit commitment in Ch. 4.

Proposition 2.12. *For any two quantum states ϱ, $\sigma \in \mathcal{B}_*(\mathcal{H})$ we have:*

$$\sup_{\omega\in\mathcal{B}_*(\mathcal{H})} \left\{ f^2(\varrho,\omega) + f^2(\sigma,\omega) \right\} = 1 + f(\varrho,\sigma)\,. \tag{2.49}$$

By Uhlmann's theorem [Uhl76, NC00], the fidelity $f(\varrho, \sigma)$ has a neat interpretation as the largest overlap of all vectors $\psi_\varrho, \psi_\sigma \in \mathcal{H} \otimes \mathcal{H}'$ that purify ϱ and σ:

Theorem 2.13. (Uhlmann's Theorem)
For any two quantum states $\varrho, \sigma \in \mathcal{B}_*(\mathcal{H})$,

$$f(\varrho, \sigma) = \max_{\psi_\varrho, \psi_\sigma} |\langle \psi_\varrho | \psi_\sigma \rangle| \,, \tag{2.50}$$

where the maximization is over all vectors $\psi_\varrho, \psi_\sigma \in \mathcal{H} \otimes \mathcal{H}'$ that purify ϱ and σ, respectively, with an ancilla space $\mathcal{H}' \simeq \mathcal{H}$.

We have seen in Sec. 2.6 that a purification is just a dilation of the quantum state, in the sense of Stinespring's representation theorem. We know from Th. 2.7 that such representations are unique up to unitary equivalence, and hence we may just as well fix two purifications $\psi_\varrho, \psi_\sigma \in \mathcal{H} \otimes \mathcal{H}'$ of the quantum states ϱ and σ, respectively, and perform the maximization over all unitary operations U on the ancilla system $\mathcal{H}'$. Uhlmann's theorem then takes the alternative form,

$$f(\varrho, \sigma) = \max_{U} |\langle \psi_\varrho | (\mathbb{1} \otimes U) \psi_\sigma \rangle| \,. \tag{2.51}$$

For pure states $\varrho = |\psi\rangle\langle\psi|$ and $\sigma = |\varphi\rangle\langle\varphi|$, the vectors coincide with their purifications, and the fidelity is then just the modulus of the inner product, $f(\psi, \varphi) = |\langle \psi | \varphi \rangle|$.

Finally, we note that for any linear operator T the operator norm $\|T\|_\infty$ equals the norm of the Schrödinger adjoint T_* on the space of trace class operators, i. e.,

$$\|T\|_\infty = \sup_{\|\varrho\|_1 \leq 1} \|T_*(\varrho)\|_1 \tag{2.52}$$

(cf. Ch. VI of [RS80] and Sec. 2.4 of [BR87] for details), which is the standard way to convert norm estimates from the Heisenberg picture into the Schrödinger picture and vice versa. For states $T_* = \varrho$, the operator norm then indeed just coincides with the trace norm, $\|T\|_\infty = \|T_*\|_1 = \|\varrho\|_1$.

2.8 Information and Entropy

In the introductory sections we have explained that we will consider physical systems as carriers of information. While we have devoted much time and space to the description of physical systems and their dynamics, up to now we have made no comment on what we mean by *information*, or how to quantify it. In everyday language, the term *information* usually refers to the content or meaning of a message. In contrast, the technical term refers to its *size*, and the possibility to store it efficiently. This is what we will cover in the present section, starting with classical systems.

From Sec. 2.1.4 we know that in the classical setting a state ϱ can be interpreted as a probability distribution $\{p_x\}_{x=1}^{|X|}$ on some finite set X. We denote the corresponding random variable by X_ϱ. By the information associated to the random variable X_ϱ we will understand the amount of uncertainty about the outcome of X_ϱ in a statistical experiment before we learn its value. The appropriate measure of information is the so-called *Shannon entropy* of X_ϱ,

$$H(X_\varrho) \equiv H(\{p_1, \cdots, p_{|X|}\}) := -\sum_{x=1}^{|X|} p_x \log p_x. \tag{2.53}$$

This definition needs some elaboration: first, by convention we set $0 \log 0 \equiv 0$. Second, the base of the logarithm is purposely left unspecified in Eq. (2.53). Fixing the base amounts to a choice of units. Throughout classical information theory the dual logarithm $\mathrm{ld} \equiv \log_2$ is very much favored; the entropy is then measured in *binary digits*, or *bits*. Although more convenient from a mathematician's point of view, natural logarithms are scarcely used. If so, the entropy is measured in *natural units*, or *nats*.

Why is the Shannon entropy a suitable measure of information? The answer is provided by Shannon's famous *noiseless channel coding theorem* [Sha48]: Assume that we are given some information source which produces a sequence $(X_i)_{i\in\mathbb{N}}$ of independent and identically distributed (i.i.d.) random variables. Shannon asked what minimal physical resources are required to store the information being produced by the source, in such a way that at some later time the information can be faithfully reconstructed — in the sense that the probability of a failure in the reconstruction procedure vanishes asymptotically as the message length increases. The answer to this question turns out to be the entropy: $H(X) \equiv H(X_1) \leq 1$ bits per symbol are required. Indeed, the Shannon entropy is a key concept in classical information theory, and plays an essential role in almost all of the basic theorems in that field. The underlying reason behind the ubiquity of entropic expressions in classical and quantum information theory is the law of large numbers, which guarantees that i.i.d. probability distributions and eigenvalues become sharply concentrated in the asymptotic limit. This will be explained in some more detail in Sec. 7.1.

There is a natural generalization of Shannon entropy to the quantum setting. It was originally introduced by John von Neumann in the study of thermodynamics and statistical mechanics, long before Shannon's ground-breaking work. The *von Neumann entropy* of a density operator $\varrho \in \mathcal{B}_*(\mathcal{H})$ is given as

$$H(\varrho) := -\mathrm{tr}\ \varrho\ \log \varrho, \tag{2.54}$$

where again the base of the logarithm is purposely left unspecified, granting a choice of units. If the eigenvalues of ϱ are given by $\{r_i\}_{i=1}^d$, we see that

$$H(\varrho) = H(\{r_i\}_{i=1}^d), \tag{2.55}$$

and hence the von Neumann entropy of the density operator ϱ is just the Shannon entropy of the spectrum of ϱ. In the classical case, where all states are completely distinguishable, von Neumann's definition and Shannon's definition are thus easily shown to coincide. For this and more properties of the entropy we refer to the text by Ohya and Petz [OP93].

While this shows that von Neumann entropy is indeed a natural generalization of Shannon entropy on formal grounds, it is not quite so clear that it is also the right definition of entropy from an operational point of view. However, this was shown by Ben Schumacher in his quantum analogue to Shannon's classical noiseless channel coding theorem [JS94, Sch95]: Given a memoryless quantum source that emits quantum states $\varrho^{\otimes n}$, we would like to encode this signal in as few qubits as possible, and send them to a receiver who will then be able to reconstruct the original state faithfully as $n \to \infty$. The maximal compression rate one can achieve in this setting is just the von Neumann entropy, $H(\varrho)$. Information spectrum methods allow a generalization of Schumacher's quantum data compression theorem to arbitrary quantum sources, as shown in Sec. 7.3.1.

Our discussion in this Section shows that entropic quantities play a major role as measures of information in both classical and quantum information theory. Some of the more important information measures that can be derived from the von Neumann entropy and appear in the theory as asymptotic rate functions for various tasks are briefly reviewed in Appendix A, alongside with their basic properties.

Chapter 3

Continuity of Stinespring's Representation

In this Chapter we will prove a continuity theorem for Stinespring's dilation: if two quantum channels are close in cb-norm, then it is always possible to find unitary implementations which are close in operator norm, with dimension-independent bounds. This result can be seen as a generalization of Uhlmann's theorem from states to channels. It allows to derive a formulation of the information-disturbance tradeoff in terms of quantum channels, as well as a continuity estimate for the no-broadcasting theorem. We briefly discuss further implications for quantum cryptography, thermalization processes, and the black hole information loss puzzle.

This Chapter represents joint work with D. Schlingemann and R. F. Werner [KSW06].

3.1 Introduction and Overview

Stinespring's dilation theorem is the basic structure theorem for quantum channels. As we have seen in Sec. 2.5, it implies that every quantum channel $T: \mathcal{B}(\mathcal{H}_B) \to \mathcal{B}(\mathcal{H}_A)$ between finite-dimensional observable algebras can be built from the basic operations of (*i*) tensoring the input with a second system in a specified state (conventionally called the *ancilla system*), (*ii*) unitary transformation on the combined input – ancilla system, and (*iii*) reduction to a subsystem. Any channel can hence be thought of as arising from a unitary evolution on a larger (*dilated*) system. The theorem even comes with a bound on the dimension of the ancilla system. Stinespring's dilation thus not only provides a neat characterization of the set of permissible quantum operations, but is also a most useful tool in quantum information science.

In this Chapter we will present a continuity theorem for Stinespring's dilation: we will show that two quantum channels, T_1 and T_2, are close in cb-norm iff we can find dilating

unitaries, V_1 and V_2, that are close in operator norm:

$$\inf_{V_1,V_2} \|V_1 - V_2\|_\infty^2 \leq \|T_1 - T_2\|_{cb} \leq 2 \inf_{V_1,V_2} \|V_1 - V_2\|_\infty \tag{3.1}$$

(cf. Sec. 2.7.1 for the interpretation of these norms).

Stinespring's representation is unique up to unitary transformations on the ancilla system. Similar to our discussion for states in Sec. 2.7.2, we may thus just as well fix two Stinespring dilations V_1 and V_2 for T_1 and T_2, respectively, and optimize over all unitaries U on the ancilla system. The continuity estimate Eq. (3.1) can then be given the alternative formulation

$$\inf_U \|(\mathbb{1}_B \otimes U)V_1 - V_2\|_\infty^2 \leq \|T_1 - T_2\|_{cb} \leq 2 \inf_U \|(\mathbb{1}_B \otimes U)V_1 - V_2\|_\infty \tag{3.2}$$

(cf. Th. 3.1 in Sec. 3.2). Hence, the continuity theorem generalizes the uniqueness clause in Stinespring's theorem to cases in which two channels T_1, T_2 differ by a finite amount. For states, i. e., channels with one-dimensional domain, dilations are usually called *purifications*, and in this special case Eq. (3.2) is an immediate consequence of Uhlmann's theorem. The proof of the continuity theorem relies on a generalization of Uhlmann's theorem from quantum states to quantum channels, and will be presented in Sec. 3.2. We initially restrict our discussion to finite-dimensional Hilbert spaces. Yet the continuity estimate Eq. (3.2) has the welcome feature of being completely independent of the dimension of the underlying Hilbert spaces, and is thus perfectly tailored for applications in which this dimension is unknown or large. In Sec. 3.6 we will briefly describe extensions of our results to infinite-dimensional systems.

We have seen in Sec. 2.5 that the ancilla system in Stinespring's representation has a natural interpretation as the environment of the physical system under investigation: the output of the channel T arises from a unitary interaction of the input state with the environment, followed by a partial trace over the degrees of freedom of the environment. Any channel T then has a *complementary channel* T_E, in which the roles of the output system and the environment are interchanged. T_E describes the information flow into the environment. Since complementary channels share a common Stinespring representation, Eq. (3.2) allows to relate the distance between two quantum channels to the distance between their complementaries. This is particularly interesting for the noiseless channel id, whose complementary channel S is completely depolarizing. The continuity theorem then entails a formulation of the information-disturbance tradeoff, which lies at the heart of quantum physics and explains why quantum information behaves so fundamentally different from its classical counterpart. We prove in Sec. 3.3 that almost all the information can be retrieved from the output of the quantum channel T by means of a decoding operation D iff T releases almost no information to the environment:

$$\frac{1}{4} \inf_D \|TD - \mathrm{id}\|_{cb}^2 \leq \|T_E - S\|_{cb} \leq 2 \inf_D \|TD - \mathrm{id}\|_{cb}^{\frac{1}{2}} \tag{3.3}$$

(cf. Th. 3.3 in Sec. 3.3). Again, no dimension-dependent factors appear in these bounds. However, we show in Sec. 3.4 that this welcome property crucially depends on the choice of the operator topology: if the cb-norm $\|\cdot\|_{cb}$ is replaced by the standard operator norm $\|\cdot\|_\infty$ in Eq. (3.3), a dimension-independent bound can in general no longer be given.

The tradeoff between information and disturbance guarantees the security of quantum key distribution in a very strong form and implies that quantum information cannot be cloned or distributed. The tradeoff theorem then amounts to a continuity estimate for the no-broadcasting theorem, presented as Cor. 3.4.

Further applications are briefly discussed in Sec. 3.5, including thermalization processes and the famous black hole information loss puzzle. In Ch. 4 we will then show how the continuity theorem allows to derive a strengthened no-go proof for quantum bit commitment.

3.2 A Continuity Theorem

In this Section we will state and prove the continuity theorem for Stinespring's representation. To set the stage, we assume two finite-dimensional Hilbert spaces $\mathcal{H}_A$ and $\mathcal{H}_B$, and quantum channels $T_1, T_2 \colon \mathcal{B}(\mathcal{H}_B) \to \mathcal{B}(\mathcal{H}_A)$ with Stinespring isometries $V_1, V_2 \colon \mathcal{H}_A \to \mathcal{H}_B \otimes \mathcal{H}_E$ as in Eq. (2.34),

$$T_1(B) := V_1^* \, (B \otimes \mathbb{1}_E) \, V_1 \,, \tag{3.4}$$

$$T_2(B) := V_2^* \, (B \otimes \mathbb{1}_E) \, V_2 \tag{3.5}$$

for all $B \in \mathcal{B}(\mathcal{H}_B)$. We can always assume that T_1 and T_2 share a common dilation space $\mathcal{H}_E$, possibly after adding some extra dimensions to one of the dilation spaces and some unitary transformations. We do not assume that either dilation $(\mathcal{H}_E, V_1)$ or $(\mathcal{H}_E, V_2)$ is minimal.

A straightforward application of the triangle inequality now immediately shows that for all $X \in \mathcal{B}(\mathbb{C}^n) \otimes \mathcal{B}(\mathcal{H}_B)$ we have

$$\begin{aligned} \|\big(\mathrm{id}_n \otimes (T_1 - T_2)\big) X\|_\infty &= \|(\mathbb{1}_n \otimes V_1^*) X (\mathbb{1}_n \otimes V_1) - (\mathbb{1}_n \otimes V_2^*) X (\mathbb{1}_n \otimes V_2)\|_\infty \\ &\leq \|\big[\mathbb{1}_n \otimes (V_1^* - V_2^*)\big] \, X \, (\mathbb{1}_n \otimes V_1)\|_\infty \\ &\qquad + \|(\mathbb{1}_n \otimes V_2^*) \, X \, \big[\mathbb{1}_n \otimes (V_1 - V_2)\big]\|_\infty \\ &\leq 2 \, \|V_1 - V_2\|_\infty \, \|X\|_\infty \,, \end{aligned} \tag{3.6}$$

independently of $n \in \mathbb{N}$, which immediately implies that

$$\|T_1 - T_2\|_{cb} \leq 2 \, \|V_1 - V_2\|_\infty \,. \tag{3.7}$$

Thus, if we can find Stinespring isometries V_1 and V_2 for the channels T_1 and T_2 which are close in operator norm, the channels will be close in cb-norm (and hence of course also in operator norm, cf. Eq. (2.43)).

As advertised in Sec. 3.1, we will now show the converse implication. Since Stinespring isometries are by no means unique, we cannot expect that *any* two given Stinespring isometries V_1, V_2 are close. The best we can hope for is that these isometries can be *chosen* close together, with dimension-independent bounds. This is in fact the essence of the continuity theorem:

Theorem 3.1. (Continuity of Stinespring's Representation)
Let $\mathcal{H}_A$ and $\mathcal{H}_B$ be finite-dimensional Hilbert spaces, and suppose that

$$T_1, T_2 \colon \mathcal{B}(\mathcal{H}_B) \to \mathcal{B}(\mathcal{H}_A) \tag{3.8}$$

are quantum channels with Stinespring isometries $V_1, V_2 \colon \mathcal{H}_A \to \mathcal{H}_B \otimes \mathcal{H}_E$ and a common dilation space $\mathcal{H}_E$. We then have:

$$\inf_U \|(\mathbb{1}_B \otimes U)V_1 - V_2\|_\infty^2 \leq \|T_1 - T_2\|_{cb} \leq 2 \inf_U \|(\mathbb{1}_B \otimes U)V_1 - V_2\|_\infty \,, \tag{3.9}$$

where the minimization is with respect to all unitary $U \in \mathcal{B}(\mathcal{H}_E)$.

If the two quantum channels coincide, $T_1 = T_2$, from the uniqueness clause in Stinespring's theorem we can conclude the existence of a unitary operation $W \in \mathcal{B}(\mathcal{H}_B \otimes \mathcal{H}_E)$ such that the intertwining relation Eq. (2.30) holds,

$$W\,(B \otimes \mathbb{1}_E) = (B \otimes \mathbb{1}_E)\,W \qquad \forall\ B \in \mathcal{B}(\mathcal{H}_B). \tag{3.10}$$

Hence, W must be decomposable as $W = \mathbb{1}_B \otimes U$ for some unitary $U \in \mathcal{B}(\mathcal{H}_E)$ acting on the ancilla system alone, and hence

$$(\mathbb{1}_B \otimes U)\,V_1 = V_2 \tag{3.11}$$

immediately follows from Eq. (2.29). We thus see from Eq. (3.9) that the continuity theorem indeed generalizes the uniqueness clause to cases in which two quantum channels differ by a finite amount, with dimension-independent bounds.

As a first step towards the proof of Th. 3.1 we will lift the equivalence Prop. 2.11 from quantum states to quantum channels. The stabilized version of the fidelity for two quantum channels T_1, T_2 has been called *operational fidelity* [BDR05]:

$$\begin{aligned} F(T_1, T_2) &:= \inf \Big\{ f(\mathrm{id} \otimes T_{1*}\,\varrho,\, \mathrm{id} \otimes T_{2*}\,\varrho) \;\mid\; \varrho \in \mathcal{B}_*(\mathcal{H}_A)^{\otimes 2}, \|\varrho\|_1 \leq 1 \Big\} \\ &= \inf \Big\{ f(\mathrm{id} \otimes T_{1*}\,|\psi\rangle\langle\psi|,\, \mathrm{id} \otimes T_{2*}\,|\psi\rangle\langle\psi|) \;\mid\; \psi \in \mathcal{H}_A^{\otimes 2}, \|\psi\| \leq 1 \Big\}, \end{aligned} \tag{3.12}$$

where minimization over pure states is sufficient by the joint concavity of the fidelity f (cf. [NC00], Th. 9.7).

Since quantum states are quantum channels with one-dimensional domain (and stabilization is not needed in this case), we have $F(\varrho, \sigma) = f(\varrho, \sigma)$ for any two quantum states $\varrho, \sigma \in \mathcal{B}_*(\mathcal{H}_A)$. The following Lemma, which we again cite from [BDR05], is then a straightforward generalization of the equivalence relation Eq. (2.48):

Lemma 3.2. (Equivalence)
For any two quantum channels $T_1, T_2 \colon \mathcal{B}(\mathcal{H}_B) \to \mathcal{B}(\mathcal{H}_A)$ *we have:*

$$1 - F(T_1, T_2) \leq \frac{1}{2} \|T_1 - T_2\|_{cb} \leq \sqrt{1 - F^2(T_1, T_2)}\,, \tag{3.13}$$

where $F(T_1, T_2)$ *denotes the* operational fidelity *introduced in Eq. (3.12).*

Proof of Lemma 3.2: The channel difference $\Phi := T_1 - T_2$ is a linear map into the $\dim \mathcal{H}_A$-dimensional system $\mathcal{B}(\mathcal{H}_A)$. Note that for such linear maps $\Phi \colon \mathcal{B} \to \mathcal{B}(\mathbb{C}^n)$, stabilization with a n-dimensional bystander system is sufficient, $\|\Phi\|_{cb} = \|\mathrm{id}_n \otimes \Phi\|_\infty$ (cf. [Pau02], Prop. 8.11). Conversion into the Schrödinger picture via the duality relation Eq. (2.52) then yields

$$\|T_1 - T_2\|_{cb} = \sup \left\{ \|\mathrm{id} \otimes (T_{1*} - T_{2*})\, \varrho\|_1 \right\} \tag{3.14}$$

where the maximization is over all $\varrho \in \mathcal{B}_*(\mathcal{H}_A \otimes \mathcal{H}_A)$ satisfying $\|\varrho\|_1 \leq 1$. The statement of the lemma now immediately follows by combining Eqs. (3.12) and (3.14) with the equivalence relation Prop. 2.11. ■

Lemma 3.2 allows us to concentrate entirely on fidelity estimates in the

Proof of Th. 3.1: As shown in Sec. 2.7.2, Uhlmann's theorem implies that the fidelity $f(\varrho, \sigma)$ of two quantum states $\varrho, \sigma \in \mathcal{B}_*(\mathcal{H}_A)$ can be given the representation

$$f(\varrho, \sigma) = \max_{U \in \mathcal{B}(\mathcal{H}_R)} |\langle \psi_\varrho | (\mathbb{1}_A \otimes U) \psi_\sigma \rangle|, \tag{3.15}$$

where ψ_ϱ and ψ_σ are any two purifications of $\varrho \in \mathcal{B}_*(\mathcal{H}_A)$ and $\sigma \in \mathcal{B}_*(\mathcal{H}_A)$, respectively, and the maximization is over all unitary operations on the dilation system $\mathcal{H}_R \simeq \mathcal{H}_A$.

Since $(\mathbb{1}_{A'} \otimes V_i)\psi$ is a purification of the output state $\mathrm{id}_{A'} \otimes T_{i*}(|\psi\rangle\langle\psi|)$, the operational fidelity $F(T_1, T_2)$ can then be expressed in terms of the Stinespring isometries V_1 and V_2 as follows:

$$\begin{aligned} F(T_1, T_2) &= \inf_\psi f(\mathrm{id}_{A'} \otimes T_{1*} |\psi\rangle\langle\psi|,\ \mathrm{id}_{A'} \otimes T_{2*} |\psi\rangle\langle\psi|) \\ &= \inf_\psi \sup_U |\langle (\mathbb{1}_{A'} \otimes V_1)\psi | (\mathbb{1}_{A'} \otimes \mathbb{1}_B \otimes U)\, (\mathbb{1}_{A'} \otimes V_2)\psi \rangle| \\ &= \inf_{\varrho \in \mathcal{B}_*(\mathcal{H}_A)} \sup_U |\mathrm{tr}\, \varrho\, V_1^* (\mathbb{1}_B \otimes U) V_2\,| \\ &= \inf_{\varrho \in \mathcal{B}_*(\mathcal{H}_A)} \sup_U \mathrm{Re}(\mathrm{tr}\, \varrho\, V_1^* (\mathbb{1}_B \otimes U) V_2)\,, \end{aligned} \tag{3.16}$$

where the maximization is now over all unitary $U \in \mathcal{B}(\mathcal{H}_E)$.

This representation is almost what we need for the desired norm estimate, since only the (fixed) Stinespring isometries V_1, V_2 and the unitary operations U on the ancilla system appear. However, from the order in which the optimization in Eq. (3.16) is performed it is clear that the optimal unitary U for the inner maximization will in general depend

on the quantum state ϱ, $U = U(\varrho)$. In order to obtain a universal unitary, observe that for fixed $\varrho \in \mathcal{B}_*(\mathcal{H}_A)$ the inner variation can be written as $\sup_U |\operatorname{tr} XU|$ with $X := \operatorname{tr}_B V_2 \varrho V_1^* \in \mathcal{B}(\mathcal{H}_E)$. It is easily seen that this supremum is attained when U is the unitary from the polar decomposition [NC00] of X, and equals $\|X\|_1$. However, since $|\operatorname{tr} XY| \leq \|X\|_1 \|Y\|_\infty$ for all $Y \in \mathcal{B}(\mathcal{H}_E)$, we can replace the supremum over all unitaries in Eq. (3.16) by a supremum over all $U \in \mathcal{B}(\mathcal{H}_E)$ such that $\|U\|_\infty \leq 1$.

With this modification both variations in Eq. (3.16) range over convex sets, and the target function is linear is both inputs. Von Neumann's minimax theorem [Neu28, Sim98] then allows us to interchange the infimum and supremum to obtain:

$$F(T_1, T_2) = \sup_{\|U\|_\infty \leq 1} \inf_{\varrho \in \mathcal{B}_*(\mathcal{H}_A)} \operatorname{Re}(\operatorname{tr} \varrho V_1^* (\mathbb{1}_B \otimes U) V_2) \,. \tag{3.17}$$

The optimization now yields a universal $U \in \mathcal{B}(\mathcal{H}_E)$. In addition, we know from our discussion above that U can always be chosen to be unitary in Eq. (3.17)[1]. Since $\|Y\|_\infty = \sup_{\|\varrho\|_1 \leq 1} |\operatorname{tr} \varrho Y|$ for any $Y \in \mathcal{B}(\mathcal{H}_A)$, we may now conclude that

$$\begin{aligned}
\inf_U \|(\mathbb{1}_B \otimes U)V_1 - V_2\|_\infty^2 &= \inf_U \left\| \left(V_1^*(\mathbb{1}_B \otimes U^* - V_2^*)\right) \left((\mathbb{1}_B \otimes U)V_1 - V_2\right) \right\|_\infty \\
&= \inf_U \sup_\varrho \operatorname{tr} \varrho \left(V_1^*(\mathbb{1}_B \otimes U^*) - V_2^*\right) \left((\mathbb{1}_B \otimes U)V_1 - V_2\right) \\
&= 2 - 2 \sup_U \inf_\varrho \operatorname{Re}(\operatorname{tr} \varrho V_1^* (\mathbb{1}_B \otimes U) V_2) \\
&= 2\left(1 - F(T_1, T_2)\right) \\
&\leq \|T_1 - T_2\|_{cb} \,,
\end{aligned} \tag{3.18}$$

where in the last step we have applied Lemma 3.2. So we have proved the left half of Eq. (3.9). ▲

The right half, which we have seen is the easier part, follows immediately from our discussion leading to Eq. (3.7) above. Alternatively, one could apply the right half of the equivalence lemma Eq. (3.13) to obtain that

$$\|T_1 - T_2\|_{cb} \leq 2\sqrt{1 - F^2(T_1, T_2)} \leq 2\sqrt{2}\sqrt{1 - F(T_1, T_2)} \,. \tag{3.19}$$

Note that without any need to invoke the minimax theorem, we can now directly conclude from Eq. (3.16) that

$$\begin{aligned}
1 - F(T_1, T_2) &\leq 1 - \sup_U \inf_\varrho \operatorname{Re} \operatorname{tr} \varrho V_1^* (\mathbb{1}_B \otimes U) V_2 \\
&= \frac{1}{2} \inf_U \|(\mathbb{1}_B \otimes U)V_1 - V_2\|_\infty^2 \,.
\end{aligned} \tag{3.20}$$

[1] *Note added in proof:* After submission of the thesis, this reasoning was found to be flawed. The optimization in Eq. (3.17) yields a universal $U \in \mathcal{B}(\mathcal{H}_E)$ which, however, may not be unitary anymore. We are only assured that $\|U\|_\infty \leq 1$. Nevertheless, by doubling the dilation space we can construct from $(\mathbb{1}_B \otimes U)\, V_2$ an isometry which dilates T_2. Th. 3.1 then holds as stated, but with an additional dimension bound on the ancilla system. The details and the corrected proof are spelled out in [KSW06].

Substituting Eq. (3.20) into Eq. (3.19), we then find

$$\|T_1 - T_2\|_{cb} \leq 2 \inf_U \|(\mathbb{1}_B \otimes U)V_1 - V_2\|_\infty \,, \tag{3.21}$$

and so we have in fact rediscovered the upper bound on the cb-norm distance. ■

3.3 Information-Disturbance Tradeoff

Since Stinespring's dilation $(\mathcal{H}_E, V)$ is essentially unique, to every quantum channel $T\colon \mathcal{B}(\mathcal{H}_B) \to \mathcal{B}(\mathcal{H}_A)$ we may associate a *complementary channel* $T_E\colon \mathcal{B}(\mathcal{H}_E) \to \mathcal{B}(\mathcal{H}_A)$, in which the roles of the output system $\mathcal{H}_B$ and the environment system $\mathcal{H}_E$ are interchanged:

$$T_E(E) := V^* (\mathbb{1}_B \otimes E)\, V \qquad \forall\ E \in \mathcal{B}(\mathcal{H}_E). \tag{3.22}$$

The channel T_E describes the information flow from the input system $\mathcal{H}_A$ to the environment $\mathcal{H}_E$. In the Schrödinger picture representation, it is obtained by tracing out the output system $\mathcal{H}_B$ instead of the ancilla system $\mathcal{H}_E$:

$$T_{E*}(\varrho) = \mathrm{tr}_B\, V \varrho V^* \qquad \Longleftrightarrow \qquad T_*(\varrho) = \mathrm{tr}_E\, V \varrho V^* \tag{3.23}$$

for all $\varrho \in \mathcal{B}_*(\mathcal{H}_A)$. Henceforth, we will usually write T_B for the channel T to better distinguish it from its complementary channel T_E.

The name *complementary channel* has been suggested by I. Devetak and P. Shor in the course of their investigation of quantum degradable channels [DS05]. Recently A. Holevo [Hol06^a] has shown that the classical channel capacity of a quantum channel T_B is additive iff the capacity of its complementary channel T_E is additive. Analogous results have been obtained independently by C. King *et al.* [KMN$^+$05] (who chose the term *conjugate channels* instead).

Since two complementary channels share a common Stinespring isometry, the continuity theorem relates the cb-norm distance between two quantum channels to the cb-norm distance between the complementary channels. The complementary channel of the noiseless channel is completely depolarizing. The continuity theorem then allows us to give a dimension-independent estimate for the information-disturbance tradeoff in terms of quantum channels:

Theorem 3.3. (Information-Disturbance Tradeoff)
Let $\mathcal{H}_A$ and $\mathcal{H}_B$ be finite-dimensional Hilbert spaces, and let $T_B\colon \mathcal{B}(\mathcal{H}_B) \to \mathcal{B}(\mathcal{H}_A)$ be a quantum channel with Stinespring dilation $(\mathcal{H}_E, V)$. By $T_E\colon \mathcal{B}(\mathcal{H}_E) \to \mathcal{B}(\mathcal{H}_A)$ we denote its complementary channel, as defined in Eq. (3.22) above. We then have the following tradeoff estimate:

$$\frac{1}{4} \inf_D \|T_B D - \mathrm{id}_A\|_{cb}^2 \leq \|T_E - S\|_{cb} \leq 2 \inf_D \|T_B D - \mathrm{id}_A\|_{cb}^{\frac{1}{2}} \,, \tag{3.24}$$

where the infimum is over all decoding channels $D\colon \mathcal{B}(\mathcal{H}_A) \to \mathcal{B}(\mathcal{H}_B)$. *In Eq. (3.24),* $S\colon \mathcal{B}(\mathcal{H}_E) \to \mathcal{B}(\mathcal{H}_A)$ *denotes a* completely depolarizing channel, *i. e.,*

$$S(E) := \mathrm{tr}(\sigma\, E)\, \mathbb{1}_A \qquad \forall\, E \in \mathcal{B}(\mathcal{H}_E) \tag{3.25}$$

for some fixed quantum state $\sigma \in \mathcal{B}_*(\mathcal{H}_E)$.

The interpretation of the tradeoff theorem is straightforward: Whenever we may find a decoding channel D such that almost all the information can be retrieved from the output of the quantum channel T_B, the norm difference $\|T_B D - \mathrm{id}_A\|_{cb}$ will be small. By the right half of Eq. (3.24), we may then conclude that the complementary channel T_E is very well approximated by a completely depolarizing channel S, and thus releases almost no information to the environment. Consequently, if a non-negligible amount of information escapes to the environment, for instance by means of a measurement performed by an eavesdropper, this will inevitably disturb the system. Hence, in quantum physics there is "no measurement without perturbation". We know from Eq. (2.43) that cb-norm and operator norm are completely equivalent in the vicinity of the noiseless channel. So any disturbance in the transmission can always be detected locally.

On the other hand, if we are assured that the channel T_E is close to some depolarizing channel S in cb-norm, the left half of Eq. (3.24) guarantees that we may find a decoding channel D which retrieves almost all the information from the B-branch of the system. Consequently, there is "no perturbation without measurement". However, in this case it is usually not enough to verify that T_E erases information locally; the channel also needs to destroy correlations. We will come back to this distinction and its implications for the interpretation of the tradeoff theorem in Sec. 3.4.

Proof of Th. 3.3: It is easily verified that a Stinespring isometry for the completely depolarizing channel $S\colon \mathcal{B}(\mathcal{H}_E) \to \mathcal{B}(\mathcal{H}_A)$, as given in Eq. (3.25), is the isometric embedding

$$V_S\colon \mathcal{H}_A \to \mathcal{H}_A \otimes \mathcal{H}_{E'} \otimes \mathcal{H}_E \qquad |\varphi\rangle \mapsto |\varphi\rangle \otimes |\psi_\sigma\rangle, \tag{3.26}$$

where $\mathcal{H}_{E'} \simeq \mathcal{H}_E$, and $|\psi_\sigma\rangle \in \mathcal{H}_{E'} \otimes \mathcal{H}_E$ is a purification of $\sigma \in \mathcal{B}_*(\mathcal{H}_E)$. Thus, the completely depolarizing channel $S \equiv S_{E'E}$ and the ideal channel id_A are indeed complementary.

The tradeoff theorem is then a straightforward consequence of the continuity theorem. Let us focus on the left half of Eq. (3.24) first, and assume that $V_T\colon \mathcal{H}_A \to \mathcal{H}_B \otimes \mathcal{H}_E$ is a Stinespring dilation for the quantum channel T_E (and its complementary channel T_B, respectively). Let $V_S\colon \mathcal{H}_A \to \mathcal{H}_A \otimes \mathcal{H}_{E'} \otimes \mathcal{H}_E$ be the Stinespring isometry of $S_{E'E}$ given by Eq. (3.26). Note that the dilation spaces $\mathcal{H}_B$ and $\mathcal{H}_A \otimes \mathcal{H}_{E'}$ are not necessarily of the same size. However, we can easily correct for this by suitably enlarging the smaller system, $\mathcal{H}_B$ say. The left half of the continuity estimate Eq. (3.9) then guarantees the

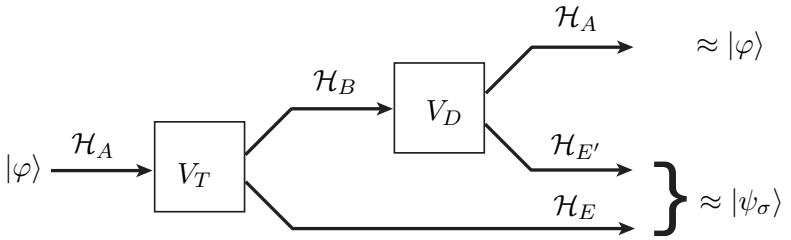

Figure 3.1: *"No perturbation without measurement." Whenever the cb-norm difference* $\|T_E - S_{E'E}\|_{cb}$ *is small, we may find a decoding channel* D_A *with Stinespring isometry* V_D *such that the concatenated isometry* $(V_D \otimes \mathbb{1}_E)V_T$ *hardly differs from the Stinespring isometry of a noiseless channel, with some fixed* $|\psi_\sigma\rangle \in \mathcal{H}_{E'} \otimes \mathcal{H}_E$.

existence of an isometry $V_D : \mathcal{H}_B \to \mathcal{H}_A \otimes \mathcal{H}_{E'}$ such that

$$\|(V_D \otimes \mathbb{1}_E)V_T - V_S\|_\infty \leq \|T_E - S_{E'E}\|_{cb}^{\frac{1}{2}}. \tag{3.27}$$

As illustrated in Fig. 3.1, the isometry V_D defines a decoding channel

$$D_A : \mathcal{B}(\mathcal{H}_A) \to \mathcal{B}(\mathcal{H}_B) \qquad D_A(A) := V_D^*(A \otimes \mathbb{1}_{E'})V_D, \tag{3.28}$$

and by the right half of the continuity estimate Eq. (3.9) we may now conclude that

$$\|T_B D_A - \mathrm{id}_A\|_{cb} \leq 2\,\|T_E - S_{E'E}\|_{cb}^{\frac{1}{2}}, \tag{3.29}$$

which proves the left half of Eq. (3.24). ▲

The proof of the right half of Eq. (3.24) proceeds very much along the same lines: Assume that $V_T : \mathcal{H}_A \to \mathcal{H}_B \otimes \mathcal{H}_E$ and $V_D : \mathcal{H}_B \to \mathcal{H}_A \otimes \mathcal{H}_{E'}$ are Stinespring isometries for the quantum channels T_B and D_A, respectively. Just as in Eq. (3.26), we again let $V_S : \mathcal{H}_A \to \mathcal{H}_A \otimes \mathcal{H}_{E'} \otimes \mathcal{H}_E$ denote the Stinespring isometry of the ideal channel id_A and its complementary channel, the completely depolarizing channel $S_{E'E}$. As before, the left half of the continuity estimate Eq. (3.9) assures us that we may find a unitary operator $U \in \mathcal{B}(\mathcal{H}_{E'E})$ such that (cf. Fig. 3.2)

$$\|(\mathbb{1}_A \otimes U)(V_D \otimes \mathbb{1}_E)V_T - V_S\|_\infty \leq \|T_B D_A - \mathrm{id}_A\|_{cb}^{\frac{1}{2}}, \tag{3.30}$$

where again we have suitably enlarged the dilation space $\mathcal{H}_{E'}$, if necessary.

Setting $\mathrm{Ad}_{V_T} := V_T^*(\cdot)V_T$ and $\mathrm{Ad}_{U^*} := U(\cdot)U^*$, we may now conclude from the right half of Eq. (3.9) that

$$\|\mathrm{Ad}_{V_T} \circ (D_{E'} \otimes \mathrm{id}_E) - S_{E'E} \circ \mathrm{Ad}_{U^*}\|_{cb} \leq 2\,\|T_B D_A - \mathrm{id}_A\|_{cb}^{\frac{1}{2}}, \tag{3.31}$$

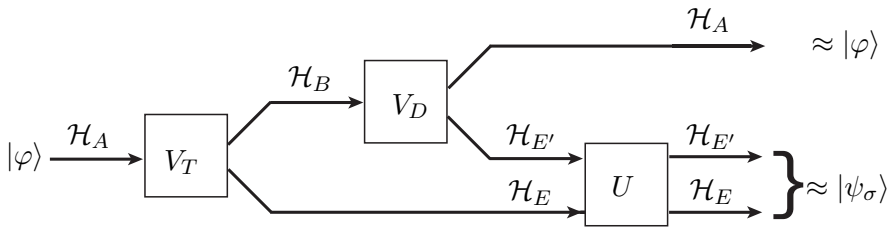

Figure 3.2: *"No measurement without perturbation." If there exists a decoding channel D_A with Stinespring isometry V_D such that the cb-norm difference $\|T_B D_A - \mathrm{id}_A\|_{cb}$ is small, we may find a unitary $U \in \mathcal{B}(\mathcal{H}_{E'}) \otimes \mathcal{B}(\mathcal{H}_E)$ such that the concatenated isometry $(\mathbb{1}_A \otimes U)(V_D \otimes \mathbb{1}_E)V_T$ is close to the Stinespring isometry of a completely depolarizing channel, with some fixed $|\psi_\sigma\rangle \in \mathcal{H}_{E'} \otimes \mathcal{H}_E$.*

which is almost the desired result. It only remains to restrict the depolarizing channel $S_{E'E}$ to the E-branch of the output system. Obviously, since $S_{E'E}$ is completely depolarizing on the combined output system $\mathcal{H}_{E'} \otimes \mathcal{H}_E$, the same holds true after a unitary rotation by U^* and the restriction to one of the branches. In particular, by setting

$$\widetilde{S}_E \colon \mathcal{B}(\mathcal{H}_E) \to \mathcal{B}(\mathcal{H}_A) \qquad \widetilde{S}_E(E) := S_{E'E} \circ \mathrm{Ad}_{U^*}(\mathbb{1}_{E'} \otimes E) \tag{3.32}$$

we obtain a completely depolarizing channel on the restricted system $\mathcal{H}_E$ such that

$$\widetilde{S}_E(E) = \mathrm{tr}(\widetilde{\sigma}\, E)\, \mathbb{1}_A \qquad \forall\ E \in \mathcal{B}(\mathcal{H}_E) \tag{3.33}$$

for $\widetilde{\sigma} := \mathrm{tr}_{E'}\, U^* |\psi_\sigma\rangle\langle\psi_\sigma| U$. It then immediately follows from Eq. (3.31) that

$$\|T_E - \widetilde{S}_E\|_{cb} \leq 2\, \|T_B D_A - \mathrm{id}_A\|_{cb}^{\frac{1}{2}}, \tag{3.34}$$

as advertised. ■

The tradeoff theorem amounts to a simple continuity estimate for the no-broadcasting and no-cloning theorems: A quantum channel $T \colon \mathcal{B}(\mathcal{H}_1) \otimes \mathcal{B}(\mathcal{H}_2) \to \mathcal{B}(\mathcal{H})$ with a triple of isomorphic Hilbert spaces $\mathcal{H}_1 \simeq \mathcal{H}_2 \simeq \mathcal{H}$ is said to *broadcast* the quantum state $\varrho \in \mathcal{B}_*(\mathcal{H})$ iff the restrictions of the output state $T_*(\varrho)$ to both subsystems coincide with the input ϱ, $\mathrm{tr}_2\, T_*(\varrho) = \varrho = \mathrm{tr}_1\, T_*(\varrho)$. The only way to broadcast a pure state $\varrho = |\psi\rangle\langle\psi|$ is to generate the product state $|\psi\rangle\langle\psi| \otimes |\psi\rangle\langle\psi|$. Thus, broadcasting of pure states is equivalent to cloning. H. Barnum *et al.* [BCF+96] have shown that a quantum channel T can broadcast two quantum states ϱ_1 and ϱ_2 iff they commute — an extension of the famous no-cloning theorem [WZ82, Die82] to mixed states. The tradeoff theorem immediately shows that approximate broadcasting is also impossible, and provides dimension-independent bounds:

Corollary 3.4. (No Broadcasting)
Let $V \colon \mathcal{H} \to \mathcal{H}_1 \otimes \mathcal{H}_2 \otimes \mathcal{H}_E$ denote a Stinespring isometry for the quantum channel

$T\colon \mathcal{B}(\mathcal{H}_1) \otimes \mathcal{B}(\mathcal{H}_2) \to \mathcal{B}(\mathcal{H})$ with local restrictions $T_1(A) := V^*(A \otimes \mathbb{1}_2 \otimes \mathbb{1}_E)V$ and $T_2(B) := V^*(\mathbb{1}_1 \otimes B \otimes \mathbb{1}_E)V$. Then there exists a completely depolarizing channel $S\colon \mathcal{B}(\mathcal{H}_2) \to \mathcal{B}(\mathcal{H})$ defined as in Eq. (3.25) such that

$$\|T_2 - S\|_{cb} \leq 2\,\|T_1 - \mathrm{id}\,\|_{cb}^{\frac{1}{2}}. \tag{3.35}$$

Hence, any broadcast channel that has reasonably high fidelity in one of the output branches releases little information to the other branch (and the environment). While Cor. 3.4 shows that neither perfect nor approximate broadcasting is possible, the bound is certainly not tight. The merit of the tradeoff theorem is a dimension-independent estimate, while optimal cloning bounds are known to depend strongly on the dimension of the underlying Hilbert space [SIG05, CF06].

3.4 Weaker Notions of Disturbance and Erasure

The tradeoff estimate established in Sec. 3.3 has the somewhat surprising and very welcome feature of being completely independent of the dimensions of the underlying Hilbert spaces, which makes the result ideally suited for applications in which these dimensions are unknown and possibly very large, as in black hole evaporation. However, this property depends crucially on the choice of the distance measure: in this Section we will give an example of a quantum channel $T\colon \mathcal{B}(\mathbb{C}^d) \to \mathcal{B}(\mathbb{C}^d)$ such that $\|T - S\|_\infty \approx 0$ for $d \to \infty$, but $\|T - S\|_{cb} \geq 1$. The example shows that operator norm and cb-norm are in general inequivalent in the vicinity of the completely depolarizing channel S. In contrast, we know from Eq. (2.43) that equivalence does hold in the neighborhood of the noiseless channel.

An example for a channel which nicely demonstrates this separation is

$$T\colon \mathcal{B}(\mathbb{C}^d) \to \mathcal{B}(\mathbb{C}^d) \qquad T := \frac{d}{d+1}S + \frac{1}{d+1}\Theta, \tag{3.36}$$

where $S\colon \mathcal{B}(\mathbb{C}^d) \to \mathcal{B}(\mathbb{C}^d)$ is the completely depolarizing channel given again by

$$S(E) = \frac{1}{d}\mathrm{tr}(E)\,\mathbb{1} \qquad \Longleftrightarrow \qquad S_*(\varrho) = \frac{\mathrm{tr}(\varrho)}{d}\,\mathbb{1}, \tag{3.37}$$

for all $E \in \mathcal{B}(\mathbb{C}^d)$ and $\varrho \in \mathcal{B}_*(\mathbb{C}^d)$, respectively. In Eq. (3.36), $\Theta\colon \mathcal{B}(\mathbb{C}^d) \to \mathcal{B}(\mathbb{C}^d)$ denotes the so-called *transpose map:* $\Theta(E) = E^t$, the matrix transpose of $E \in \mathcal{B}(\mathbb{C}^d)$. While Θ is linear, unital and positive, it is not completely positive, and thus cannot be implemented as a quantum channel [Pau02]. However, we will show in Prop. 3.5 below that the convex mixture T nonetheless remains a valid quantum channel.

Noting that $\|\Theta\|_\infty = 1$ and $\|\Theta\|_{cb} = d$ [Pau02], it then immediately follows that

$$\|T - S\|_\infty = \frac{1}{d+1}\,\|\Theta - S\|_\infty \leq \frac{1}{d+1}\left(\|\Theta\|_\infty + \|S\|_\infty\right) = \frac{2}{d+1}, \tag{3.38}$$

and thus $\lim_{n\to\infty} \|T - S\|_\infty = 0$. On the other hand, making again use of the triangle inequality we have the lower bound

$$\|T - S\|_{cb} = \frac{1}{d+1} \|\Theta - S\|_{cb} \geq \frac{1}{d+1} \left(\|\Theta\|_{cb} - \|S\|_{cb} \right) = \frac{d-1}{d+1}. \tag{3.39}$$

This demonstrates the suggested separation between cb-norm and operator norm as $d \to \infty$. It only remains to show that T is a quantum channel, which will become clear from the proof of

Proposition 3.5. *Let $\Theta\colon \mathcal{B}(\mathbb{C}^d) \to \mathcal{B}(\mathbb{C}^d)$ be the transpose map, and $S\colon \mathcal{B}(\mathbb{C}^d) \to \mathcal{B}(\mathbb{C}^d)$ the completely depolarizing channel as given in Eq. (3.37). Then*

$$T_p\colon \mathcal{B}(\mathbb{C}^d) \to \mathcal{B}(\mathbb{C}^d) \qquad T_p := (1-p)\, S + p\, \Theta \tag{3.40}$$

for $p \in [0,1]$ defines a quantum channel iff $p \leq \frac{1}{d+1}$.

Proof: While T_p is clearly linear, unital and positive for all $p \in [0,1]$, it is not necessarily completely positive, since the transpose map Θ does not have this property. In order to test for complete positivity, it is sufficient to apply the Schrödinger dual T_{p*} to half of a maximally entangled state $|\Omega\rangle := \frac{1}{\sqrt{d}} \sum_{i=1}^d |i,i\rangle$ on $\mathbb{C}^d \otimes \mathbb{C}^d$. In fact, it follows from Jamiolkowski's duality theorem (cf. [Jam72] and Th. 2.3.4 in [Key02]) that a linear map $R\colon \mathcal{B}(\mathbb{C}^d) \to \mathcal{B}(\mathbb{C}^d)$ is completely positive iff $\varrho := R_* \otimes \mathrm{id}\, |\Omega\rangle\langle\Omega|$ is a quantum state.

We will now apply this statement to the family T_p. It is easily seen from Eq. (2.21) that the Schrödinger dual Θ_* coincides with Θ, $\Theta_* = \Theta$. Straightforward calculation shows that

$$\varrho_p := \left(T_{p*} \otimes \mathrm{id} \right) |\Omega\rangle\langle\Omega| = \frac{1-p}{d^2}\, \mathbb{1} \otimes \mathbb{1} + \frac{p}{d}\, \mathbb{F}\,, \tag{3.41}$$

where $\mathbb{F} := \sum_{i,j} |i,j\rangle\langle j,i|$ is the so-called *flip operator*. Note that $\mathbb{F}^* = \mathbb{F}$ and $\mathbb{F}^2 = \mathbb{1}$. Quantum states of the form

$$\varrho = \alpha\, \mathbb{1} + \beta\, \mathbb{F} \tag{3.42}$$

are usually called *Werner states* [Wer89]. In order to see for which values α and β the operator ϱ describes a quantum state, it is useful to rewrite Eq. (3.42) in terms of the eigenprojections $P_\pm$ of the flip operator $\mathbb{F}$, i. e., $\mathbb{F}\, P_\pm\, |\psi\rangle = \pm\, P_\pm |\psi\rangle$ with

$$P_+ := \frac{\mathbb{1} + \mathbb{F}}{2} \qquad \text{and} \qquad P_- := \frac{\mathbb{1} - \mathbb{F}}{2}. \tag{3.43}$$

P_+ is the projection onto the symmetric (Bose) subspace, while P_- describes the projection onto the antisymmetric (Fermi) subspace. Observing that $P_+ + P_- = \mathbb{1}$ and $P_+ - P_- = \mathbb{F}$ and substituting these expressions into Eq. (3.42), we see that

$$\varrho = (\alpha + \beta)\, P_+ + (\alpha - \beta)\, P_-, \tag{3.44}$$

which is positive iff $\alpha \geq \beta$. This implies that the output state ϱ_p, as given in Eq. (3.41), is a quantum state (and thus T_p is completely positive, by the Jamiolkowski duality) iff $p \leq \frac{1}{d+1}$, as suggested. ■

Hayden *et al.* [HLS+04] have recently proven that random selections of unitary matrices generically show an even stronger separation:

Proposition 3.6. *Let $\varepsilon, \delta > 0$. Then for sufficiently large d there is a pair of channels $R, S : \mathcal{B}(\mathbb{C}^d) \to \mathcal{B}(\mathbb{C}^d)$ such that $\|R - S\|_\infty \leq \varepsilon$ and $\|R - S\|_{cb} \geq 2 - \delta$.*

Proof: In the terminology of Ref. [HLS+04], a quantum channel $R: \mathcal{B}(\mathbb{C}^d) \to \mathcal{B}(\mathbb{C}^d)$ is called *ε-randomizing* iff

$$\|R_*(\varrho) - S_*(\varrho)\|_1 \leq \varepsilon \qquad \forall \ \varrho \in \mathcal{B}_*(\mathbb{C}^d)\,, \tag{3.45}$$

where S again denotes the completely depolarizing channel, as in Eq. (3.37) above. Eq. (3.45) implies the norm estimate $\|R - S\|_\infty \leq \varepsilon$, as required in Prop. 3.6.

Hayden *et al.* show that for $d > \frac{10}{\varepsilon}$, such an ϵ-randomizing quantum channel can be obtained with high probability from a random selection of at most $\mu := \lceil \frac{134}{\varepsilon^2} d \operatorname{ld} d \rceil$ unitary operators $\{U_i\}_{i=1}^{\mu} \subset \mathcal{B}(\mathbb{C}^d)$,

$$R(E) := \frac{1}{\mu} \sum_{i=1}^{\mu} U_i^* \, E \, U_i \,. \tag{3.46}$$

In striking contrast, exact randomization of quantum states (such that $\varepsilon = 0$ in Eq. (3.45)) is known to require an ancilla system of dimension $d^2 \gg \mu$ [AMT+00]. However, the significant reduction in the size of the ancilla space comes at a price: while the randomizing map R erases local information, it preserves almost all the correlations with a bystander system if d is sufficiently large. In fact, it is straightforward to show the upper bound

$$\|(R_* - S_*) \otimes \mathrm{id}\, |\Omega\rangle\langle\Omega|\|_1 \geq 2 - \frac{2\mu}{d^2}\,, \tag{3.47}$$

and the same holds true for any other channel R with an ancilla system of dimension $o(n^2)$. Eq. (3.47) implies the bound $\|R - S\|_{cb} \geq 2 - \delta$, where $\delta := \frac{2\mu}{d^2}$ can be made as small as desired by choosing d sufficiently large. ■

From the right half of the tradeoff theorem Eq. (3.24) we may then conclude that for none of these channels T and R, as defined in Eqs. (3.36) and (3.46), will it be possible to find a decoding channel D such that the randomized information can be recovered from the ancilla system alone. Information may remain hidden in quantum correlations and cannot be retrieved locally. Since standard operator norm and cb-norm coincide for channels with classical (Abelian) output space (cf. Th. 3.9 in Paulsen's text [Pau02]), this separation demonstrates a truly quantum-mechanical effect.

While these examples illustrate that it is in general not possible to upper bound the cb-norm $\|\cdot\|_{cb}$ in terms of the operator norm $\|\cdot\|_\infty$ with a dimension-independent estimate, a dimension-*dependent* bound can of course be given. In fact, for any linear map $R\colon \mathcal{A} \to \mathcal{B}(\mathbb{C}^d)$ with an arbitrary (possibly infinite-dimensional) C^*-algebra $\mathcal{A}$ we have $\|R\|_{cb} \leq d\,\|R\|_\infty$ [Pau02]. The transpose map Θ shows that this bound can be tight.

3.5 Further Applications

Th. 3.3 provides a dimension-independent tradeoff estimate in terms of stabilized operator norms. Christandl and Winter [CW05] have recently obtained complementary entropic bounds: Assume that a uniform quantum ensemble $E_1 := \{\frac{1}{d}, |i\rangle\}$ of basis states of the Hilbert space $\mathcal{H} \simeq \mathbb{C}^d$ and the Fourier-rotated ensemble $E_2 := \{\frac{1}{d}, U|i\rangle\}$ have both nearly maximal Holevo information when sent through the quantum channel $T_B\colon \mathcal{B}(\mathbb{C}^d) \to \mathcal{B}(\mathbb{C}^d)$:

$$\chi\big(T_B(E_k)\big) \geq \operatorname{ld} d - \varepsilon \tag{3.48}$$

for $k = 1, 2$ and some (small) $\varepsilon > 0$, where $\chi(T_B(E_1))$ denotes the Holevo information of the output ensemble $T_B(E_1) := \{\frac{1}{d}, T_{B*}(|i\rangle\langle i|)\}$, and analogously for the rotated ensemble $T_B(E_2) := \{\frac{1}{d}, T_{B*}(U|i\rangle\langle i|U^*)\}$ (cf. Appendix A). Christandl and Winter can then conclude from their entropic uncertainty relation that the *coherent information*

$$I_c\Big(T_B, \frac{1}{d}\mathbb{1}\Big) = H\bigg(\frac{1}{d}\sum_{i=1}^{d} T_{B*}\big(|i\rangle\langle i|\big)\bigg) - H\Big(T_{B*} \otimes \mathrm{id}\,\big(|\Omega\rangle\langle\Omega|\big)\Big) \geq \operatorname{ld} d - 2\varepsilon \tag{3.49}$$

is likewise large, where again $|\Omega\rangle = \frac{1}{\sqrt{d}}\sum_{i=1}^{d} |i, i\rangle$ is a maximally entangled state on $\mathbb{C}^d \otimes \mathbb{C}^d$. As a consequence, there exists a decoding operation $D\colon \mathcal{B}(\mathbb{C}^d) \to \mathcal{B}(\mathbb{C}^d)$ such that $F_c(T_B\, D) \geq 1 - 2\sqrt{2\varepsilon}$ [SW02], where $F_c(R)$ denotes the *channel fidelity* of the quantum channel R, as defined in Sec. 2.7.1.

The faithful transmission of the maximally entangled state $|\Omega\rangle$ is not sufficient to conclude that $T_B D \approx \mathrm{id}$ in operator norm with dimension-independent bounds [KW04]. But as explained in some more detail in Sec. 5.2.2, it is possibly to find a subspace $\mathcal{H}' \subset \mathcal{H}$ with $\dim \mathcal{H}' \geq \frac{1}{2}\dim \mathcal{H}$ such that $\|T_B' D' - \mathrm{id}\,'\|_{cb} \leq 13\,\varepsilon^{\frac{1}{8}}$ [BKN00, KW04], where T_B' and D' are restricted to $\mathcal{H}'$. The tradeoff theorem then guarantees that $\|T_E' - S\|_{cb} \leq 8\,\varepsilon^{\frac{1}{16}}$ for some completely depolarizing channel S. Thus, in combination with Th. 3.3 the existence of highly reliable detectors for a basis and its conjugate alone imply a stabilized version of privacy, which is much stronger than the entropic version that appears in [CW05]. The improvement comes at the expense of a smaller code space. However, for many cryptographic applications this is an exponentially large space, hence its reduction by a factor 1/2 does not affect the rate of the protocol.

The information-disturbance tradeoff also plays the central role in the infamous black hole information loss puzzle: black holes emit thermal Hawking radiation [Haw74],

which contains (almost) no information about the previously absorbed quantum states. We can model this evaporation process as an (almost) completely depolarizing quantum channel, $T_E \approx S$. The tradeoff theorem then suggests that all the data about the formation of the black hole resides inside the event horizon, and could at least in principle be retrieved from there. However, the black hole may eventually evaporate completely, seemingly erasing all this information in the process and hence violating the unitarity of quantum mechanics.

The tradeoff theorem provides the explicit bounds for the quantum estimate. Our results also show that for large objects with many internal degrees of freedom — such as the entirety of all the information swallowed by a black hole —, the estimate crucially depends on the choice of the operator topology. If only an unstabilized estimate of the form $\|T_E - S\|_\infty \leq \varepsilon$ can be guaranteed, information may be hidden in quantum correlations between the thermal radiation and the black hole final state.

Similar conclusions apply to thermalization processes, in which a quantum system approaches an equilibrium state via repeated interaction with an environment. In so-called *collision models* [SZS$^+$02, ZSB$^+$02], the evolution of the thermalizing quantum system is described in terms of a quantum channel T_E. If T_E is almost completely depolarizing in cb-norm, all the information about the initial state of the system will have dissipated into the environment, and can at least in principle be retrieved from there.

Braunstein and Pati [BP06] have explored the consequences of the information-disturbance tradeoff for the physics of black holes and thermalization in greater detail.

3.6 Summary and Conclusions

In conclusion, in this Chapter we have shown and explored a continuity theorem for Stinespring's dilation theorem: two quantum channels, T_1, T_2 are close in cb-norm iff there exist corresponding Stinespring isometries, V_1 and V_2, which are close in operator norm. When applied to the noiseless channel $T_1 = \mathrm{id}$, the continuity theorem yields a formulation of the information-disturbance tradeoff in which both information gain and disturbance are measured in terms of operator norms, complementing recently obtained entropic bounds. In the form we have presented it, the continuity theorem applies to quantum channels on finite-dimensional quantum systems and yields dimension-independent bounds. This makes the result ideally suited for applications to situations in which these dimensions are large or possibly unknown.

The absence of dimension-dependent factors in the continuity bounds Eq. (3.9) seems to indicate that the result is not restricted to the finite-dimensional setting. Extensions of the continuity theorem to completely positive maps between arbitrary C^*-algebras are currently under investigation.

We have already seen in Sec. 2.1.3 that infinite dimensions lead to a number of complications, since not all states on infinite-dimensional systems can be represented as density operators. Generalizing the definition of normal states given in Sec. 2.1.3, we say that a quantum channel $T\colon \mathcal{B}(\mathcal{H}_A) \to \mathcal{B}(\mathcal{H}_B)$ is *normal* iff $\lim_{n\to\infty} T(A_n) = T(A)$ for every sequence $(A_n)_{n\in\mathbb{N}}$ of norm-bounded increasing operators with least upper bound $A \in \mathcal{B}(\mathcal{H}_A)$. The normal channels T are then precisely those for which the duality relation Eq. (2.21) continues to hold (cf. [Dav76], Ch. 9). Non-normal (or *singular*) channels do not have a Schrödinger dual.

However, as long as the Hilbert spaces are separable and all systems obey generic energy constraints, the state space will be compact [Hol03], the channels respecting these energy constraints will be normal, and our proof of the continuity theorem and the tradeoff bounds then goes through unchanged. In particular, compactness of the state space guarantees that von Neumann's minimax theorem can be applied [Sim98]. Thus, all the results presented in the preceding sections continue to hold in this practically relevant setting.

Chapter 4

Quantum Bit Commitment: A Strengthened No-Go Theorem

Bit commitment is a cryptographic primitive involving two mistrustful parties, conventionally called Alice and Bob. Alice is supposed to submit an encoded bit of information to Bob in such a way that Bob has (almost) no chance to identify the bit before Alice later decodes it for him, whereas Alice has (almost) no way of changing the value of the bit once she has submitted it: in technical terms, a good bit commitment protocol should be simultaneously *concealing* and *binding*.

Bit commitment protocols whose security is based on the laws of quantum physics alone are generally held to be impossible. In this Chapter we will give a strengthened and explicit proof of this result. We extend its scope to a much larger variety of protocols, which may have an arbitrary number of rounds, in which both classical and quantum information is exchanged, and which may include aborts and resets. Moreover, we do not consider the receiver, Bob, to be bound to a fixed "honest" strategy, so that "anonymous state protocols", which were recently suggested as a possible way to beat the known no-go results are also covered. We show that any concealing protocol allows Alice to find a cheating strategy which is universal in the sense that it works against any of Bob's strategies. Moreover, if the concealing property holds only approximately, the cheat goes undetected with a high probability, which we explicitly estimate. The proof is based on the continuity theorem for Stinespring's dilation, as presented in Ch. 3, and uses an explicit formalization of general two party protocols, possibly of independent interest. The result also provides a natural characterization of protocols that fall outside the standard setting of unlimited available technology, and thus may allow secure bit commitment. We present a new such protocol whose security, perhaps surprisingly, relies on decoherence in Bob's lab.

This work was done in collaboration with G. M. D'Ariano, D. Schlingemann, and R. F. Werner [DKS+06].

4.1 Introduction

Let us begin this Section with an example due to Winter *et al.* [WNI03], which nicely illustrates how bit commitment naturally arises in everyday life: Suppose that chess masters Alice and Bob are playing the world championship. As night falls in and people start to feel hungry, Alice and Bob realize that they will have to interrupt the game and resume it on the next morning. We assume that it is Alice's turn when this happens. Now there is a problem: if Alice plays her turn before they both leave and go home, Bob will have the entire night to think of his response, giving him an unfair advantage. But if Alice postpones her move, then she will have the entire night to come up with a clever move, and Bob would find that highly unsatisfactory.

A bit commitment scheme can provide a fair solution. In such a scheme, Alice commits to a piece of information during a commitment phase. In our example, this would be the specification of her next move, but for simplicity we will always assume in the sequel that her message can be represented by a single bit. At some later point in time, e.g. on the following morning, Alice will unveil the message by sending some opening information to Bob during an opening phase.

Such a protocol is said to be *concealing* if the information sent by Alice during the commitment phase is (almost) perfectly hidden from Bob. The protocol is said to be *binding* if Alice cannot submit a certain message in the commitment phase and then successfully reveal a different message in the opening phase. A bit commitment protocol is secure iff it is both concealing and binding.

Bit commitment has immediate practical applications (for chess masters Alice and Bob), and is also known to be a very powerful cryptographic primitive. It was conceived by Blum [Blu83] as a building block for secure coin tossing. Bit commitment also allows to implement secure oblivious transfer [BBC$^+$91, Cre94, Yao95], which in turn is sufficient to establish secure two-party computation [Kil88, CVT95].

A trusted third party makes bit commitment a trivial task: in our example, Alice would write down her move on a piece of paper and hand it over to the referee before they leave. The referee would pass on the information to Bob on the next morning, who could then check that Alice indeed plays the move she previously committed to.

But Alice and Bob know that trusted third parties are hardly ever available when both glory and money are at stake. If they don't trust the referee, an alternative solution could be for Alice to write down her move on a piece of paper, which is then locked in a safe and sent to Bob, whereas Alice keeps the key. On the next morning, she will unveil by handing over the key to Bob, so that he can check Alice's move against his record. However, Bob has a well-equipped toolbox at home and may have been able to open the safe in the meantime. So while this scheme may offer reasonably good practical security, it is in principle insecure. Yet all bit commitment schemes that have wide currency today rely on such technological constraints: not on strongboxes and keys,

but on unproven assumptions that certain computations are hard to perform. Several such protocols have been suggested, either computationally binding [Blu83, BCC88, Hal95, HM96] or computationally concealing [Nao91, OVY92]. Cryptographers have long known that without such technological constraints, bit commitment (like any other interesting two-party cryptographic primitive) cannot be securely implemented in a classical world [Kil88].

It has therefore been a long-time challenge for quantum cryptographers to find *unconditionally secure* quantum bit commitment protocols, in which — very much in parallel to quantum key distribution [BB84, Eke91] — security is guaranteed by the laws of quantum physics alone.

4.1.1 Quantum Bit Commitment and the No-Go Theorem

The first quantum bit commitment protocol is due to Bennett and Brassard and appears in their famous 1984 quantum cryptography paper [BB84], in a version adapted to coin tossing. In their scheme, Alice commits to a bit value by preparing a sequence of photons in either of two mutually unbiased bases, in a way that the resulting quantum states are indistinguishable to Bob. The authors show that their protocol is secure against so-called *passive cheating*, in which Alice initially commits to the bit value k, and then tries to unveil $1 - k$ later. However, they also prove that Alice can cheat with a more sophisticated strategy, in which she initially prepares pairs of maximally entangled states instead, keeps one particle of each pair in her lab and sends the second particle to Bob. It is a direct consequence of the EPR effect [EPR35] that Alice can then unveil either bit at the opening stage by measuring her particles in the appropriate basis, and Bob has no way to detect the difference.

Subsequent proposals for bit commitment schemes tried to evade this type of attack by forcing the players to carry out measurements and communicate classically as they go through the protocol. At a 1993 conference Brassard, Crépeau, Jozsa, and Langlois presented a bit commitment protocol [BCJ$^+$93] that was claimed and generally accepted to be unconditionally secure.

Three years later it was then realized by Lo and Chau [LC97, LC98], and independently by Mayers [May96, May97, BCM$^+$97] that all previously proposed bit commitment protocols are vulnerable to a generalized version of the EPR attack that renders the BB84 proposal insecure — a result which they slightly extended to cover quantum bit commitment protocols in general. In essence, their proof goes as follows: At the end of the commitment phase, Bob will hold one out of two quantum states ϱ_k as proof of Alice's commitment to the bit value $k \in \{0, 1\}$. Alice holds its purification ψ_k, which she will later pass on to Bob to unveil. For the protocol to be concealing, the two states ϱ_k should be (almost) indistinguishable, $\varrho_0 \approx \varrho_1$. But as we have seen in Sec. 2.7.2, Uhlmann's theorem [Uhl76, NC00] then implies the existence of a unitary

transformation U that (nearly) rotates the purification of ϱ_0 into the purification of ϱ_1. Since U is localized on the purifying system only, which is entirely under Alice's control, Lo-Chau-Mayers argue that Alice can switch at will back and forth between the two states, and is not in any way bound to her commitment. As a consequence, any concealing bit commitment protocol is argued to be necessarily non-binding.

These results still hold true when both players are restricted by superselection rules [KMP04]. So while the proposed quantum bit commitment protocols may offer good practical security on the grounds that Alice's EPR attack is hard to perform with current technology, none of them is unconditionally secure. Spekkens and Rudolph [SR01] have extended the no-go theorem by providing explicit bounds on the degree of concealment and bindingness that can be achieved simultaneously in any bit commitment protocol, some of which they showed can be saturated.

4.1.2 Two Camps

In view of these negative results, subsequent research has primarily focused on bit commitment under plausible technological constraints, such as a limited classical [CCM98, DHR$^+$04] or quantum [DFS$^+$05] memory, or the difficulty of performing collective measurements [Sal98]. In an alternative approach, researchers have slightly modified the standard setting to evade the no-go theorem: Kent [Ken99, Ken05] has shown that relativistic signalling constraints may facilitate secure bit commitment when Alice and Bob each run two labs a (large) distance apart and security is maintained through a continual exchange of messages. A different variant was introduced by Hardy and Kent [HK04], and independently by Aharonov *et al.* [ATV$^+$00]: in *cheat-sensitive* bit commitment protocols, both players may have the chance to cheat, but face the risk of their fraud being detected by the adversary. Building on Kent's original proposal [Ken03], the tradeoff between bindingness and concealment in quantum *string* commitment protocols has recently been investigated [BCH$^+$05, Jai05].

At the same time, the Lo-Chau-Mayers no-go theorem [LC97, May97] is continually being challenged. Yuen and others have repeatedly expressed doubts in Mayer's opaque paper [May97], arguing that the no-go proof is not general enough to exclude all conceivable quantum bit commitment protocols. Several protocols have been proposed and claimed to circumvent the no-go theorem (see [Yue00, Yue04, Yue05, Che01] and references therein, as well as D'Ariano's account [Dar02^a, Dar02^b] of the controversy). These protocols seek to strengthen Bob's position with the help of 'secret parameters' or 'anonymous states', so that Alice lacks some information to cheat successfully: while Uhlmann's theorem would still imply the existence of a unitary cheating transformation as described above, this transformation might be unknown to Alice.

Two camps seem to have formed, a large one comprising most of the community, in which the impossibility of quantum bit commitment is accepted on the basis of the

Lo-Chau-Mayers proof, and a smaller group of sceptics, which is not convinced, even though no provably secure protocol, and hence a counterexample to the no-go result, has surfaced so far.

It appears that much of this controversy stems from slightly differing approaches to the problem. A good way to pinpoint the basic disagreement is Kerckhoffs' principle [Ker83], which goes back to the 19th century military cryptographer Auguste Kerckhoffs and is now universally embraced by cryptographers [TW06, Rud02]. The principle states that the security of a cryptographic protocol should not rely on keeping parts of the algorithm secret. In the words of Bruce Schneier, "every secret creates a potential failure point. Secrecy, in other words, is a prime cause of brittleness — and therefore something likely to make a system prone to catastrophic collapse" [Sch02]. In this respect every secret parameter chosen by a human in a cryptographic protocol — e. g. a password — is regarded as a potential weakness. For this reason cryptographers usually think of their algorithms as being executed by machines, whose blueprints can be published without jeopardizing the security of the system.

Anonymous states and other secret parameters used in Yuen's protocols are apparently regarded as a violation of Kerckhoffs' principle, which suggests a restriction to fixed and automatizable strategies for both players. Deviations from these strategies are considered an attempted fraud. The Kerckhoffian security analysis then does not hold any provisions for the case in which *both* parties deviate from their 'honest' strategies. Therefore Lo-Chau-Mayers only consider the final committed state given that Bob sticks to his publicly known strategy, since Alice's cheat only has to work against this strategy. So while Kerckhoffs' principle is certainly high on the list of desiderata for cryptographic protocols, it appears that Lo-Chau-Mayers only show that there is no bit commitment protocol *satisfying Kerckhoffs' principle*, whereas the next best thing, e.g., an anonymous state protocol might still exist.

Another possible origin for disagreement is the style of Mayer's paper [May97], along the lines of Mark Kac's dictum, "A demonstration convinces a reasonable man; a proof convinces a stubborn man" [Kac68]. In this sense, i.e., according to the standards of "stubborn" mathematics or mathematical physics, Mayers gives merely a demonstration. Since the argument against Kerckhoffian protocols only involves the state directly after commitment, Mayers declares it irrelevant to formalize the class of two-party protocols, even though an insufficiently specified domain usually leaves a no-go "theorem" rather fuzzy. Other aspects of the problem (e.g., the use of classical and quantum information together) get a similarly rough treatment. This may be a symptom of the "Four Page Pest", i.e., the disease of cramming an argument onto four pages in PRL format, although its shortest intelligible presentation requires more than six. In any case, it appeared to us high time to convince ourselves, and hopefully some other stubborn men, of the exact scope and status of the no bit commitment statements.

4.1.3 A Stronger No-Go Theorem: Overview and Outline

In this Chapter we propose to resolve the bit commitment controversy with a strengthened no-go theorem. We will give a precise description of general two-party protocols, which we hope no longer shows the hard work of keeping it fully explicit but still notationally manageable. This description should also be helpful for analyzing protocols for other tasks, involving any number of parties. Our description of bit commitment does not assume Kerckhoffs' principle, so that Bob is not honor bound to a particular course of action. Nevertheless, we show that any concealing protocol allows Alice a universal cheating strategy, working against all strategies of Bob simultaneously. Moreover, our result is stable against small errors, in the sense that nearly concealing protocols allow a nearly perfect cheat, with explicit universal error bounds. The result is based on the continuity theorem for Stinespring's representation, as presented in Ch. 3.

Our proof applies to bit commitment protocols with any (finite) number of rounds during each the commitment, holding, and opening phase. It includes a full treatment of the classical and quantum information flow and also covers aborts and resets. The proof is not restricted to quantum systems that can be described in finite-dimensional Hilbert spaces, but can be extended to continuous-variable systems with generic (global) energy constraints. The strengthened no-go theorem shows the insecurity of all recently proposed bit commitment protocols [Yue00, Yue04, Yue05, Che01]. A preliminary version of the proof, restricted to single-round commitments, has appeared in [BDR05]. Our results generalize that of Ozawa [Oza01] and recent work by Cheung [Che05], who showed that Alice can still cheat in protocols with secret parameters for the simpler case of perfect concealment, and without a full reduction. Cheung's estimates [Che06] for approximately concealing protocols depend on the dimension of the underlying Hilbert spaces, and hence cannot rule out bit commitment protocols with high-dimensional systems.

We also classify those protocols that fall outside the standard setting, and thus may allow secure bit commitment. Inspired by Yuen's *anonymous state* idea, we propose a new such bit commitment protocol whose security — perhaps paradoxically — relies on decoherence in Bob's lab. Interestingly, this protocol explores a purely quantum-mechanical effect: the distinction between the local erasure of information and the destruction of quantum correlations, as explained in Sec. 3.4. Well-known classical bit commitment protocols whose security relies on noisy communication channels are briefly reviewed, too.

This Chapter is organized as follows: In Sec. 4.2 we give a detailed description of the setup for quantum bit commitment protocols, and list important types of protocols that fall within our definition. This will serve to specify the domain for the proof of the strengthened no-go theorem, which is then presented in Sec. 4.3. Sec. 4.4 investigates provably secure bit commitment protocols whose security is built on decoherence in either Alice's or Bob's lab, or in the transmission line. In Sec. 4.5 we briefly describe

how to extend the no-go theorem to quantum bit commitment protocols in infinite-dimensional Hilbert spaces with global energy constraints. Appendix B contains the necessary background on direct sums and quantum-classical hybrid systems.

4.2 The Setup

In this Section we describe the task of quantum bit commitment, and define what a successful bit commitment protocol would have to achieve. We have attempted not to exclude any possibilities, and have avoided all simplifications "without loss of generality" at this stage. In this way we hope to separate, more clearly than our predecessors, the definition of bit commitment to which the statement "Bit commitment is impossible" refers and, on the other hand, the simplifications which we will make in the course of the proof of this statement.

The analysis will be based solely on the principles of quantum mechanics, including classical physics. We do not consider relativistic signalling constraints, which are known to facilitate secure bit commitment [Ken99, Ken05]. We impose as a *finiteness condition*, that all classical messages can only take finitely many values, that all quantum systems can be described in a finite-dimensional Hilbert space, and that the total number of messages exchanged must be uniformly bounded.

4.2.1 Description in Plain English

The Basic Task — Bit Commitment is a cryptographic primitive involving two mistrustful parties, conventionally called Alice and Bob. Alice is supposed to submit an encoded bit of information to Bob in such a way that Bob has (almost) no chance to identify the bit before Alice decodes it for him, and Alice has (almost) no way of changing the value of the bit after she has submitted it. In other words, Bob is interested in *binding* Alice to some commitment, whereas Alice would like to *conceal* her commitment from Bob.

Protocols and Strategies — A *protocol* first of all regulates the exchange of messages between Alice and Bob, such that at every stage it is clear what type of message is expected from the participants, although, of course, their content is not fixed. The expected message types can be either classical or quantum or a combination thereof, with the number of distinguishable classical signals and the dimension of the Hilbert spaces fixed. The type of messages can depend on classical information generated previously. The collection of all these instructions will be called the *communication interface* of the protocol.

A particular plan for operating a local laboratory to supply the required messages, is called a *strategy*. A strategy could determine that some message sent is obtained

from a measurement on a system available in the local lab, but it could also specify the arbitrary invention of a classical value to be sent and the fresh preparation of an accompanying quantum system. We typically denote Alice's strategy by a and Bob's by b.

The second key element of the protocol specifies definite procedures for Alice to follow if she wants to commit the bit values 0 or 1, respectively. These special *honest strategies* will be denoted by a_0 and a_1.

Phases of the Protocol — In any commitment scheme, we can distinguish three phases. The first is the *commitment phase*, with a possibly complicated exchange of classical and quantum messages. By definition, at the end of this phase, the bit value is considered to be committed to Bob but, supposedly, concealed from him.

Alice and Bob then might split up for a while, without further communication. In this *holding phase* typically only local operations are possible, i.e., Bob might attempt to read the committed bit, and Alice might attempt to prepare a cheat.

Finally they get in touch again to open the commitment. In the *opening phase*, Alice sends to Bob the value of the bit she claims to have committed, together with all quantum or classical information needed for Bob to check this claim against his own (classical and quantum) records. Bob then performs a suitable *verification measurement* with two possible outcomes: either an "ok" confirming Alice's claim, or a "not ok". The protocol might also be ended in a *public opening*, which requires Alice and Bob to meet, bringing with them all quantum and classical systems in their possession, explaining what strategies they were using, and allowing Bob to choose arbitrary measurements on all these systems to verify, with Alice staying on to watch. That is, no possibility of cheating, withholding information, or making false claims about the outcome of the verification exists in a public opening.

Conditions on Successful Protocols — We assume that Alice's strategies a_0 and a_1 can be distinguished with high probability by Bob's verification measurement (which depends on the value of the bit Alice claims to have committed). So if he tests for the bit value "0", and Alice honestly played a_0 he will get an "ok" with probability $\geq (1-\eta)$ for some (small) $\eta \geq 0$, but if Alice played a_1 the verification will give "ok" only with probability $\leq \eta$, and similarly with tests for "1". In this case we call the protocol *η-verifiable*, or *η-sound*. Since this condition depends only on honest strategies, and one pair of measurements, it is very easy to satisfy.

We call a protocol *ε-concealing*, if Alice's honest strategies cannot be distinguished by Bob (up to an error ε). In general, of course, the probabilities he measures while applying his protocol b depend on whether Alice chooses a_0 or a_1. Here we require that no matter what strategy b Bob uses and no matter what measurement he makes, these probabilities never differ by more than ε throughout the commitment and holding phase. Note that the concealing condition makes no statement whatsoever about other

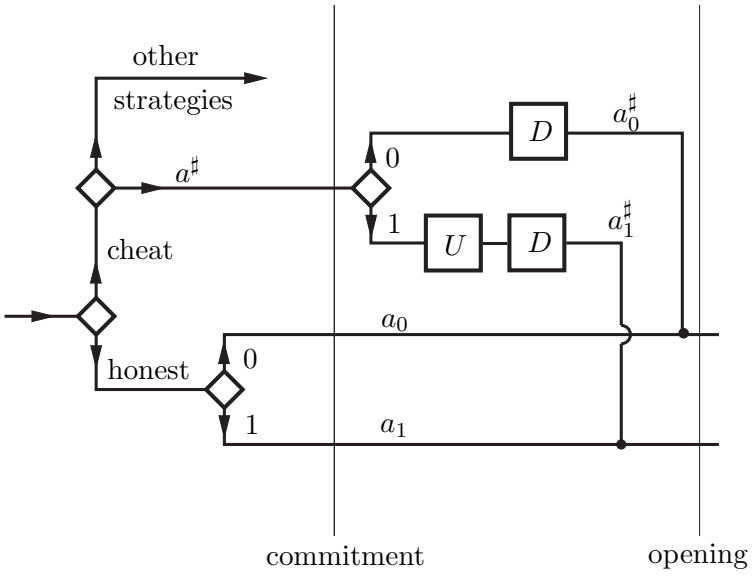

Figure 4.1: *Alice's basic strategic choices. Decisions she must take are indicated by diamonds, some actions necessary for a typical cheating strategy by squares. The cheating strategies $a_0^\sharp$ and $a_1^\sharp$ are identical throughout the commitment phase, and might be equal to a purification of the honest strategy a_0. Then U indicates a unitary cheating transformation, and D the introduction of suitable decoherence to reverse the purification. In the opening phase the cheating strategies are identical to their honest counterparts.*

strategies of Alice. If Alice cheats, there is usually nothing to be concealed anyway.

A *δ-cheating strategy* for Alice is a pair of strategies $a_0^\sharp$ and $a_1^\sharp$ such that Bob cannot distinguish a_0 from $a_0^\sharp$, and a_1 from $a_1^\sharp$ better than with a probability difference δ, at any time, including the opening phase. Of course, these conditions would be trivially satisfied for $a_0 = a_0^\sharp$ and $a_1 = a_1^\sharp$. What makes $(a_0^\sharp, a_1^\sharp)$ cheating strategies is that Alice does not actually make the decision about the value of the bit until after the commitment phase. That is, the strategies $a_0^\sharp$ and $a_1^\sharp$ must be the same throughout the commitment phase, and can only differ by local operations carried out in the holding or opening phase. Note, however, that Alice might have to decide from the outset that she wants to cheat, since the strategies $a_i^\sharp$ might be quite different from both a_0 and a_1. Fig. 4.1 illustrates Alice's basic choices as she goes through the protocol. If no δ-cheating strategy exists for Alice, we call the protocol δ-binding.

The condition we impose here is much stronger than the condition that Bob's standard verification measurements are fooled by the cheat (perhaps with a bound on the success probability): we require that no measurement whatsoever could detect a difference. With a public opening rule, one could even say that after the cheat not even Alice herself could help Bob to tell the difference. Clearly, these conditions make it very

hard for Alice to cheat. Therefore, our proof that Alice can still cheat under such conditions automatically includes all protocols with weaker conditions on successful cheats.

Real Time Checks for Cheating — It is perhaps helpful to point out the difference between two kinds of checks on Alice's honesty, which Bob might perform. We have granted him unlimited technological power in the definition of ε-concealing protocols. But for running the protocol no such fantastic abilities are required, and he will not actually do all those complicated tests. In fact, the concealing and binding properties of the protocol cannot be ascertained by any practical tests, but are there to be checked theoretically by Alice and Bob on the basis of the publicly available description of the protocol. It is on the basis of such considerations that Alice and Bob will consent to use the protocol in the first place.

During a single run of the protocol, Bob can employ some tests on Alice's behavior as part of the protocol. If Bob suspects a problem he may be entitled to calling an abort of the protocol (clearly a classical message), and the procedure would start at the beginning. The total number of such resets must be limited on the grounds of bounding Alice's probability of cheating. The possibility of such checks at run-time are the main reason why we must consider protocols with a large number of rounds, possibly differing from run to run.

Result — We will prove in Sec. 4.3.7 that any protocol which is ε-concealing allows a δ-cheating strategy for Alice, where $\delta \leq 2\sqrt{\varepsilon}$. These bounds coincide with those obtained by Spekkens and Rudolph [SR01] in the Kerckhoffian setting.

As illustrated in Fig. 4.1, Alice's cheating strategy $a^\sharp$ consists in playing a purification of the honest strategy a_0 throughout the commitment and holding phase. If she then opts for the bit value $k = 1$ instead, she will apply a unitary operation U on the purifying system, and then follow the honest strategy a_1 from there.

4.2.2 Formal Description of Protocols

In this Section we will cast the above description more explicitly into the formalism of quantum theory. Thereby we further reduce possible ambiguities in the statement of the problem, but also prepare the notation for the proof.

We will generally identify systems by their observable algebras. This has the advantage that combinations of classical and quantum information are naturally covered: a quantum system with Hilbert space $\mathcal{H}$ is then represented by the algebra $\mathcal{B}(\mathcal{H})$ of operators on $\mathcal{H}$, and a system characterized by a classical parameter x, and has Hilbert space $\mathcal{H}_x$ in that case is described by the direct sum $\bigoplus_x \mathcal{B}(\mathcal{H}_x)$. A state on such an algebra is of the form $\bigoplus_x p_x\rho_x$, and is specified first by a probability distribution $\{p_x\}_x$ for the x's, and second by a collection of density operators ρ_x on $\mathcal{H}_x$, which are used to

compute expectations if the value of the classical parameter is known to be x. Since this formalism for handling classical information in protocols is not generally familiar, we describe it in some more detail in Appendix B.

Many algebras (indexed by the nodes of the communication tree) will appear in the description of the protocol, indicating that with each operation the type of quantum system in the respective lab might change completely. By choosing the lab algebras large enough this dependence might be avoided. However, even when the lab systems remain the same, it is helpful to keep the distinguishing indices for keeping track of the progress of the protocol.

4.2.2.1 The communication tree

At every stage of the protocol a certain amount of shared classical information will have accumulated. Classical information never gets lost, so the stages of the protocol, together with the currently available classical information naturally form the nodes of a tree, which we call the *communication tree*. An example is depicted in Fig. 4.2. Every node x carries the following information:

1. Whose turn is it: Alice's or Bob's? This follows from the position of the node in the tree, when we assume without loss of generality, that Bob always starts, and from then turns alternate.

2. What are the classical signals, which might be sent from this person to the other? The admissible signals form a finite set M_x by assumption. This set labels the branches going from this node to successor nodes, which we denote by $x' = xm$, for $m \in M_x$.

3. For each possible classical signal, what kind of quantum system is accompanying it? If the classical message is m, we take its observable algebra to be $\mathcal{M}_m^x$, and assume this to be the full algebra of $(d \times d)$-matrices for some $d = d(x, m) < \infty$. The value $d(x, m) = 1$ ($\simeq$ no accompanying quantum system) is a possible choice.

4. Each node x is completely characterized by the entire history of the classical messages exchanged between Alice and Bob, i.e., we can write $x = m_1 m_2 \cdots m_N$.

At every node, we denote the observable algebras of Alice's and Bob's laboratories by $\mathcal{A}_x$ and $\mathcal{B}_x$, respectively. These are only partly determined by the communication interface, and depend on the strategy, which we sometimes emphasize by writing $\mathcal{A}_x(a)$ and $\mathcal{B}_x(b)$. The description of the communication step below shows in detail how these algebras develop as one moves along the communication tree. Let X_c denote the set of nodes at which a commitment is supposed to be reached. Since only local operations and the opening phase follow, we can consider these as the leaves of the communication

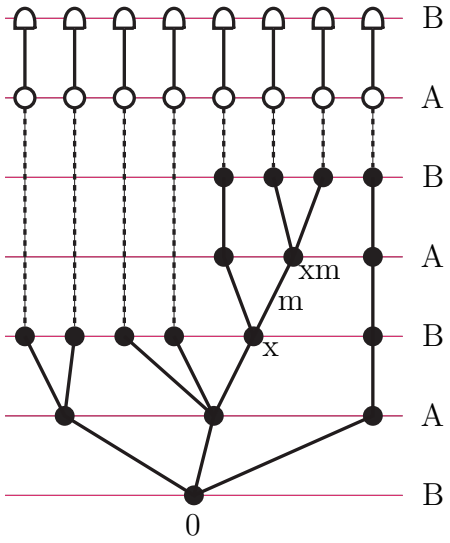

Figure 4.2: *Example of a communication tree. Each node corresponds to one history of classical communications, with the different lines from each node representing a possible classical signal. The dashed lines represent the holding phase, in which no communication occurs, followed by the opening move (open circle) by Alice, and a measurement by Bob.*

tree. The joint observable algebra at that stage is

$$\bigoplus_{x \in X_c} \mathcal{A}_x(a) \otimes \mathcal{B}_x(b) \tag{4.1}$$

(cf. Appendix B for the interpretation of direct sums). The algebras of Alice and Bob could themselves be direct sums, representing classical information only available to Alice and Bob, respectively, but we do not look at this for the moment.

4.2.2.2 The elementary communication step

Now consider some node x, and assume that it is Alice's turn (everything holds mutatis mutandis for Bob). We know that some message $m \in M_x$ is expected from Alice, accompanied by a quantum system with observable algebra $\mathcal{M}_m^x$. The most general way of doing this is a quantum operation sending states on $\mathcal{A}_x$ to states on $\bigoplus_m \mathcal{A}_{xm} \otimes \mathcal{M}_m^x$. Written in the Heisenberg picture, Alice hence chooses a channel

$$T_x(a) \quad : \quad \bigoplus_{m \in M_x} \mathcal{A}_{xm}(a) \otimes \mathcal{M}_m^x \to \mathcal{A}_x(a) \tag{4.2}$$

$$\mathcal{B}_{xm}(b) \quad = \quad \mathcal{M}_m^x \otimes \mathcal{B}_x(b). \tag{4.3}$$

Here we have added a parameter a to T_x, to make it clear that choosing these channels for all x is precisely what defines Alice's strategy. Note that the choice of the channel

includes that of their domain and co-domain algebras. The channel $T_x(a)$, together with the input state determines the probabilities for the classical outcomes m. Of course, the channel could be one that simply forces one of the results. Hence m could equally well be the result of Alice's free choice of strategy, or of a measurement on a system she recently obtained from Bob. If m is found, Alice also splits the output system into a part $\mathcal{A}_{xm}(a)$ which Alice keeps, and the part $\mathcal{M}_m^x$ she sends to Bob. This splitting is included in the specification of $T_x(a)$. That $\mathcal{M}_m^x$ changes ownership is expressed in the above equation by including it in Bob's algebra at the next round (i.e., $\mathcal{B}_{xm}$) as a tensor factor. At Bob's nodes everything is the same, but since we always order tensor factors as Alice $\otimes$ Message $\otimes$ Bob, the analogues of the above equations at Bob's nodes are

$$T_x(b) \quad : \quad \bigoplus_{m\in M_x} \mathcal{M}_m^x \otimes \mathcal{B}_{xm}(b) \to \mathcal{B}_x(b) \tag{4.4}$$

$$\mathcal{A}_{xm}(a) \quad = \quad \mathcal{A}_x(a) \otimes \mathcal{M}_m^x. \tag{4.5}$$

Recall that initially there is neither shared classical information nor quantum systems, so for the node 0 ="start" we have $\mathcal{A}_0 = \mathcal{B}_0 = \mathbb{C}$. Therefore, the state at the commitment stage

$$\rho_c(a,b) : \bigoplus_{x\in X_c} \mathcal{A}_x(a) \otimes \mathcal{B}_x(b) \ \to \mathbb{C} \tag{4.6}$$

is completely determined by the choice of protocols a, b. Similarly, the final state, on which Bob carries out the verification measurement will be denoted by $\rho_f(a,b)$.

4.2.2.3 Can Bob distinguish Alice's strategies?

In the concealment condition, as well as in the description of cheating strategies, it is important to decide whether Bob can distinguish two strategies of Alice at commitment time. Clearly, this depends only on the restriction $\rho_c^B(a_i,b)$ of the state $\rho_c(a_i,b)$ to Bob's laboratory, which has observable algebra $\bigoplus_x \mathcal{B}_x$.

The security criterion given in Sec. 4.2.1 asks for the largest probability difference obtainable by Bob. We know from Sec. 2.7.2 that it is convenient to express this as a trace norm difference: the largest difference of expectations in "yes-no" experiments with density matrices ρ_1, ρ_2 is $\sup_F |\mathrm{tr}(\rho_1 - \rho_2)F|$, where F ranges over all effects F with $0 \le F \le \mathbb{1}$. That is, the largest probability difference is $\frac{1}{2}\|\rho_1 - \rho_2\|_1$, where $\|\cdot\|_1$ denotes the trace norm. This naturally leads us to the following definitions of concealing protocols and cheating strategies:

Definition 4.1. (Concealment)
We say that a protocol is ε-concealing *iff for all strategies* b *of Bob*

$$\|\rho_c^B(a_0,b) - \rho_c^B(a_1,b)\|_1 \le 2\,\varepsilon. \tag{4.7}$$

When this condition holds with $\varepsilon = 0$*, we say that the protocol is* perfectly concealing.

Definition 4.2. (Bindingness)
A δ-cheating *strategy* $(a_0^\sharp, a_1^\sharp)$ *is characterized by the inequality*

$$\|\rho_f^B(a_i^\sharp, b) - \rho_f^B(a_i, b)\|_1 \leq 2\,\delta, \tag{4.8}$$

for $i = 0, 1$ *and all* b*. If no* δ*-cheating strategy exists, the protocol is called* δ-binding.

Def. 4.2 requires a cheating strategy to work against *all* of Bob's strategies — not only against some *fixed* strategy, as suggested by Kerckhoffs' principle. We will show in Sec. 4.3.7 that Alice can always find such a universally good cheating strategy. As explained in Sec. 4.1.3, this extends the no-go theorem to protocols relying on secret parameters or "anonymous states".

Note that one possible measuring strategy for Bob is to actually make the measurement at some earlier time, record the result, and send only dummy messages to Alice afterwards. So saying that two strategies are ε-equivalent *at* some stage is the same as saying that they are equivalent *up to* that stage of the protocol. Hence, the ε-concealing condition implies the only apparently stronger statement that at no time during the commitment phase Bob is able to discriminate the honest commitments better than with probability ε.

4.2.3 Protocols Covered by our Definition

In this Section we describe some ideas from the literature about possible protocols, in increasing complexity. Of course, none of them are ultimately successful. But this is in many cases not obvious from the outset, so these ideas serve well to illustrate the richness of two-party protocols as formalized in our scheme.

4.2.3.1 The beginning

As explained in Sec. 4.1.1, the first observation concerning quantum bit commitment was made in the classic paper of Bennett and Brassard on quantum cryptography [BB84]. In this basic scenario the commitment phase has only one round, in which Alice prepares one of two orthogonal Bell states $\psi_0, \psi_1 \in \mathcal{H}_A \otimes \mathcal{H}_B$. These have the same restriction on Bob's system, so the protocol is perfectly concealing. But they are also connected by a unitary on Alice's side (as all maximally entangled states are), and this unitary constitutes her *sneak flip* cheating strategy, which under these circumstances also works perfectly.

4.2.3.2 Alice sends a state

The natural generalization of this protocol is to replace the Bell states by arbitrary pure states generated by Alice [LC97, May97]. When these have the same restriction

on Bob's side, they are purifications of the same state, and hence are connected by a partial isometry on Alice's side, which serves as a sneak flip operation. A crucial step is now to go away from perfect concealment ($\varepsilon = 0$ in Eq. (4.7)), which seems to have been considered first in [May97]. In this case one has to use a continuity result for purifications, i.e., that nearby states have nearby purifications. In other words, one needs an estimate of Uhlmann's fidelity (which measures the distance between purifying vectors), and the trace norm, as provided in Sec. 2.7.2.

4.2.3.3 Classical communication

Classical communication occurs naturally in cryptographic protocols, so it needs to be included in the analysis. In contrast to some of our predecessors, who choose a purely quantum description from the outset, we treat classical information explicitly throughout. In particular, classical information in the Lo-Chau-Mayers approach is treated quantum-mechanically and sent over noiseless quantum channels, while our description explicitly allows information transfer over classical channels, and thus provides a natural setting to include purely classical protocols in the analysis.

Cheating becomes harder for Alice if the protocol requires some exchange of classical information, for she no longer has full control over the purification spaces of the two commitment states. Roughly speaking, unitaries which introduce superpositions of states, which belong to different classical values already sent to Bob, are forbidden. In the formalism introduced above this means that Alice has to find a cheating unitary for every classical communication history x.

Mayers' heuristic paper [May97] has some provisions for this case, by sending classical values to a special quantum repository in the environment, and effectively coherentifying all classical information. In our work, the classical communication flow emerges naturally as a framework for the description of the protocol. This approach should also prove helpful in the analysis of other cryptographic tasks.

4.2.3.4 Bob supplies the paper

The protocols so far were characterized by the property that Bob really had no strategic choices to make during the commitment phase. Hence the state at the end of the commitment phase, written in our scheme as $\rho_c(a, b)$, really does not depend on Bob's strategy b. So Alice only has to connect the purifications of two states which are explicitly known to her. Clearly, her task of finding a clever sneak flip becomes harder if there is a proper dependence on b. Lo-Chau-Mayers restrict their analysis to those protocols in which Bob follows a specified 'honest' strategy $b_\star$, which is assumed to be publicly known in accordance with Kerckhoffs' principle. In these cases, Alice knows how to cheat, and the no-go result immediately applies.

As explained in Sec. 4.1.3, we do not require that Bob follows such a publicly known standard strategy. Alice then indeed has to find a sneak flip working for all of Bob's admissible strategies b. The easiest such protocol begins with Bob sending a system to Alice, in some state known only to him (in [Yue05] this is called an *anonymous* state). The honest strategies require Alice to encode the bit by using this system in some way and then returning a committing system to Bob. Effectively Alice now chooses not a *state* but a *channel* to encode her commitment. The purification idea and Uhlmann fidelity estimate no longer work for this, so these protocols are not covered by Lo-Chau-Mayers. Instead, the purification construction has to be generalized to the Stinespring representation of channels, and an appropriate continuity result has to be shown. This will be done in Sec. 4.3.

4.2.3.5 A decoherence monster in Bob's lab

That the idea of states supplied by Bob may introduce interesting new aspects is demonstrated by a scenario which is not a bit commitment protocol in the sense of this Section, because it makes additional assumptions about things happening in Bob's lab: Suppose that after Bob has sent some quantum state to Alice, a decoherence monster (such as the cleaning service) enters his lab, and all quantum information is destroyed. Only his classical records survive. That is, he still knows what preparation he made, but cannot use the entangled records he made during the preparation. Now suppose that Alice and Bob can rely on this happening. Then they can design a bit commitment protocol that works. So, paradoxically, the monster strengthens Bob's position, because it weakens the assumptions about his ability to break the concealment. Hence one can make protocols which are binding in the strong sense described above, but concealing only if we assume that coherence in Bob's lab is indeed destroyed. We will analyze this possibility in Sec. 4.4.2.

4.2.3.6 Alice can choose more strategies

An apparent generalization would allow Alice to choose her honest strategy a_0 at will from some set $\mathsf{A_0}$, and a_1 from $\mathsf{A_1}$. The idea is that now some $a_0 \in \mathsf{A_0}$ might well be distinguishable from some $a_1 \in \mathsf{A_1}$ for Bob. Concealment under such circumstances means that Bob, on seeing data compatible with some a_0 during the commitment or holding phase can never be sure that they do not come from a certain a_1. In other words, for every $a_0 \in \mathsf{A_0}$ there must be an ε-equivalent strategy $a_1 \in \mathsf{A_1}$. But then, according to our result, Alice can develop a sneak flip attack on the basis of these two protocols alone.

4.2.3.7 More communication in the holding phase

In Sec. 4.2.1 we have excluded any communication in the holding phase, and, apart from a single message from Alice to Bob, also in the opening phase. There is, however, no problem in allowing such communication, and some protocols, like Kent's protocol using relativistic signal speed constraints [Ken99, Ken05], require a lot of communication in the holding phase.

Of course, protocols with no rounds at all in the holding phase are directly covered by our definition. The only strategic difference between holding and commitment phase is that Alice's cheating strategies $a_0^\sharp$ and $a_1^\sharp$ are only required to coincide during the commitment phase. She might start cheating with different tricks for 0 and 1 during the holding phase.

Clearly, declaring the holding phase a part of the commitment phase only weakens Alice's cheating possibility. However, she does not need these extra options anyway: a sneak flip attack at the end of the holding phase is always possible, as we show.

4.2.3.8 Aborts and resets

Often in cryptography one considers protocols which allow the parties to call an "abort". We can distinguish two kinds of abort: when a *constructive abort*, or *reset* occurs, the protocol is started anew, whereas at a *full abort* the whole protocol is terminated as unsuccessful.

Both kinds of aborts are covered in our scheme, but they would be typical of different phases. Resets are quite natural in the commitment phase. For example, Bob might make a test measurement on some message he receives, and refuse to continue if there is a slight deviation from what is expected from Alice playing honest. A reasonable requirement at this point is that the probability for reaching a commitment after some number of rounds with an honest Alice is positive. Then allowing even more retrials one could bring the probability for reaching a commitment close to one, and allow some arbitrary choice in the remaining cases, i.e., if the allotted total number of rounds is exhausted without a commitment. In this way one would get a protocol satisfying our finiteness condition, while retaining the potential value of resets for a commitment protocol. Strictly speaking, resets can only occur during the commitment phase, since we have demanded a partitioning of each protocol run into three successive phases (without relapses into earlier phases). However, the holding phase can be essentially united with the commitment phase (see Sec. 4.2.3.7). Hence we can effectively also cover constructive aborts during the holding phase.

In the opening phase we can consider *full*, or destructive aborts. This is a move right to an endpoint of the communication tree, labelled accordingly. Clearly this possibility weakens Bob's discrimination powers, and makes it much easier for Alice to cheat. In

particular, each sneak flip attack becomes successful. Therefore, the abort possibility does not seem to present any interesting strategic options for quantum bit commitment.

4.3 Proof

In the exposition of the task of bit commitment and the admissible protocols we have tried not to restrict generality by simplifying assumptions, in order not to weaken the scope of the no-go theorem. This leads to a rather wild class of strategies to be considered: arbitrarily many rounds of communication of varying length, infinite-dimensional local lab Hilbert spaces, and all that. Clearly, in the course of the proof we want to get rid of this generality. The main idea for simplifications is that obviously inferior methods of analysis for Bob, or inferior cheating methods for Alice need not be considered. We therefore begin with an explanation of what it means that one strategy is "obviously inferior", or *weaker* than another (cf. Sec. 4.3.1).

The first application of this idea is the process of *purification*, by which a general strategy is turned into another one, which avoids all measurements not demanded by the communication interface, and turns all decohering operations into coherent information transfer to ancillas. Stinespring's dilation theorem guarantees that this can always be done. The purifications result in *locally coherent strategies*, which will be crucial for Alice's cheat later on, and have been a part of all no-go results.

Once a player has chosen a locally coherent strategy, it is possible to reduce the lab spaces considerably. For example, if a strategy requires the choice of a mixed state, this state may have an infinite-dimensional support Hilbert space. Its purification, however, is a single vector, so up to a unitary transformation, which can be absorbed into subsequent operations, it suffices to take a one-dimensional Hilbert space. We show that this works for operations as well: for every locally coherent strategy there is a stronger one (in the sense of Sec. 4.3.1), using only finite-dimensional Hilbert spaces, with a universal dimension bound depending only on the dimension of quantum messages exchanged so far. In particular, an infinite-dimensional lab space will not give more power to Bob. This will be shown in Sec. 4.3.3, and leads to the consequence that effectively (up to any desired level of accuracy) we need only consider a finite number of strategies for Bob.

The next step is in some sense a dual of purification: purification means that we can avoid measurements during a protocol, deferring all such operations to the final measurement. Similarly, we can move the acts of decision making during the protocol to the very beginning, by introducing a *strategy register* (see Sec. 4.3.4), which is described in the Hilbert space $\ell^2(S)$, for some finite set S of strategies. The choice of a strategy is then expressed by preparing some initial state of the strategy register, and then letting controlled unitaries transcribe this information into suitable operations at all later rounds. Let us denote by b_σ Bob's strategy of installing the strategy register

mechanism, and preparing the initial state σ for that register. The state $\rho_c(a, b_\sigma)$ at commitment time then depends linearly on σ, and after tracing out Alice's lab, we find a channel $\Gamma^B(a)$ depending on Alice's strategy a, such that

$$\Gamma^B(a) \quad : \quad \bigoplus_{x \in X_c} \mathcal{B}_x \to \mathcal{B}(\ell^2(S)) \tag{4.9}$$

$$\text{tr } \sigma \, \Gamma^B(a)(B) \quad = \quad \text{tr } \rho_c^B(a, b_\sigma) \, B \tag{4.10}$$

for all $B \in \oplus_x \mathcal{B}_x$. This channel now summarizes everything that Bob can possibly learn about Alice's strategy by choosing his own strategy and making a measurement in his lab after the commitment. In a simple, purely Kerckhoffian scenario the analogous object is just the state at commitment time, since one does not allow Bob a choice of different legitimate strategies. However, in our more general framework we do need to consider the dependence on σ, and correspondingly cheats which work uniformly well for all σ.

As an instructive special case, we next suppose that the protocol is *perfectly* concealing, which is expressed by $\Gamma^B(a_0) = \Gamma^B(a_1)$. We show in Sec. 4.3.5 that Alice then has a perfect cheat. Its existence is guaranteed by the uniqueness clause in the Stinespring dilation theorem. From this prototype of Alice's cheat one can see how an approximate cheat in response to approximate concealment $\Gamma^B(a_0) \approx \Gamma^B(a_1)$ should work.

In Sec. 4.3.6 we look more carefully into the kind of approximation $\Gamma^B(a_0) \approx \Gamma^B(a_1)$ sufficient to draw the desired conclusion. It turns out that we need to consider a special attempt of concealment breaking for Bob, namely keeping an entangled record of the strategy register and making a joint measurement on the rest of his system and this "backup copy" after commitment. Clearly, this is a legitimate attempt in our framework, and hence must already be implicit in the strategies controlled by the strategy register. However, making this scheme explicit provides the right kind of norm (cb-norm) on channels so that a small $||\Gamma^B(a_0) - \Gamma^B(a_1)||_{cb}$ guarantees the existence of an approximately ideal cheat. The technical result guaranteeing this is the continuity theorem for the Stinespring dilation construction, as presented in Ch. 3.

4.3.1 Comparing the Strength of Strategies

Consider two strategies a and a' of Alice. We will say that a' is *stronger* than a, if whatever Alice can achieve by strategy a she can also achieve by a'. More explicitly, we require that there is a *revert* operation $R_x : \mathcal{A}_x(a) \to \mathcal{A}_x(a')$ bringing Alice back to strategy a at whatever node x she so chooses (observe the direction of arrows due to the Heisenberg picture). That she actually comes back to a is guaranteed inductively,

i.e., we require that

$$R_x T_x(a) = T_x(a') \bigoplus_{m \in M_x} (R_{xm} \otimes \mathrm{id}_{\mathcal{M}_m^x}) \tag{4.11}$$

at Alice's nodes and

$$R_{xm} = R_x \otimes \mathrm{id}_{\mathcal{M}_m^x} \tag{4.12}$$

at Bob's nodes.

Tracing this all the way back to the root of the communication tree we get, for any of Bob's protocols b, and for any stage of the protocol, in particular for the commitment stage X_c,

$$\mathrm{tr}\ \rho_c(a,b)\Big(\bigoplus_x F_x \otimes G_x\Big) = \mathrm{tr}\ \rho_c(a',b)\Big(\bigoplus_x R_x(F_x) \otimes G_x\Big). \tag{4.13}$$

Taking $F_x = \mathbb{1}_x$ in Eq. (4.13) (corresponding to the partial trace over Alice's lab space in the Schrödinger picture) we see that Bob's subsystems are completely unaffected, i.e., Bob will never be able to tell the difference between a and a'. The strategic significance of passing to a stronger strategy is different for Alice and for Bob.

For Bob a stronger b' is just another strategy to be considered in the concealing condition and in the condition for a successful cheat. Since Bob does not loose any discriminating power in playing coherent, Alice (and we) might as well assume that he is always using the strongest strategy available. This simplifies the analysis, as we will see in more detail below.

For an honest Alice there is no option. Whatever the honest strategies a_0 and a_1 specify, she has to follow. However, since Bob will never know the difference, it is easy to check from the definitions of concealment and bindingness in Sec. 4.2.2.3 that whenever (a_0, a_1) is a bit commitment protocol with security parameters ε and δ, then so is any pair of stronger strategies (a_0', a_1'), with the same parameters. Hence we could assume for the sake of an impossibility proof that Alice's honest strategies are strengthened in some way. However, there is hardly an advantage in that assumption, and we will not do so.

For a cheating Alice, using all the power of her infinitely well equipped lab, and hence using the strongest available strategies is clearly the best choice. Indeed, this will be the only difference between the honest and the cheating strategies during the commitment phase: these consist of playing until commitment a particular strengthening of an honest strategy, namely the local purification discussed in the next subsection.

4.3.2 Local Purification

Intuitively, maintaining coherence during quantum operations is more demanding than allowing thermal noise and other sources of decoherence to have their way. Therefore, doing only those measurements needed for satisfying the communication interface rules,

but avoiding all other decoherence should lead to a stronger protocol in the sense of Sec. 4.3.1. Such simplified "locally coherent" strategies are more easily expressed in terms of operators acting on Hilbert spaces than in terms of superoperators acting on algebras. Therefore we need a notation for the message Hilbert spaces as well, i.e., we set $\mathcal{M}_m^x = \mathcal{B}(\mathcal{K}_m^x)$, where $\dim \mathcal{K}_m^x = d(x,m)$ is the dimension parameter from the description of the communication tree in Sec. 4.2.2.1.

Definition 4.3. (Locally Coherent Strategy)
A strategy a (of Alice) is called locally coherent *iff for all communication nodes x we have $\mathcal{A}_x(a) = \mathcal{B}(\mathcal{H}_x(a))$ and, at all of Alice's nodes, the respective quantum channel $T_x(a) : \bigoplus_m \mathcal{A}_{xm}(a) \otimes \mathcal{M}_m^x \to \mathcal{A}_x(a)$ from Eq. (4.2) is given by operators*

$$V_{x,m}(a) : \mathcal{H}_x(a) \to \mathcal{H}_{xm}(a) \otimes \mathcal{K}_m^x \tag{4.14}$$

such that

$$T_x(a)\Big(\bigoplus_m A_m \otimes Y_m\Big) = \sum_m V_{x,m}(a)^*\big(A_m \otimes Y_m\big)V_{x,m}(a) \tag{4.15}$$

for all $A_m \in \mathcal{B}(\mathcal{H}_{xm}(a))$ and $Y_m \in \mathcal{B}(\mathcal{K}_m^x)$.

The point here is that each summand in this $T_x(a)$ is *pure*, i.e., given by a single Kraus operator $V_{x,m}(a)$. This is equivalent to the property that the m^{th} term in this sum cannot be decomposed into a non-trivial sum of other completely positive maps, which would in turn correspond to the extraction of further classical information. Using a non-pure map in a strategy would therefore mean to exercise less than the maximal control allowed by quantum theory. Note that m is in general a random outcome, but Alice can make it deterministic by choosing her strategy a corresponding to $V_{x,m}(a) = \delta_{m,m_0}V_x$, with an isometry V_x.

We have seen in Sec. 2.5 that Stinespring's dilation theorem provides the canonical way to convert any strategy into a locally coherent one. We will use Stinespring's dilation Th. 2.6 several times, but ignore the uniqueness statement Th. 2.7 for the moment. Then we can iteratively generate a locally coherent protocol $\breve{a}$ from a, together with the required revert operations showing that $\breve{a}$ is indeed stronger than a. Suppose the space $\mathcal{H}_x(\breve{a})$ and the revert channel $R_x : \mathcal{A}_x(a) \to \mathcal{B}(\mathcal{H}(\breve{a}))$ has already been defined along with these objects for all earlier nodes. We need to extend this definition to all successor nodes xm. If the node x belongs to Bob, there is nothing to do since Eq. (4.12) explicitly defines R_{xm}. At Alice's nodes, we apply Stinespring's theorem to the composition

$$R_xT_x(a) : \bigoplus_{m\in M_x} \mathcal{A}_{xm}(a) \otimes \mathcal{M}_m^x \ \to \mathcal{B}(\mathcal{H}_x(\breve{a})). \tag{4.16}$$

The dilation theorem then yields a representation π_x of $\bigoplus_{m\in M_x} \mathcal{A}_{xm}(a) \otimes \mathcal{M}_m^x$ on some Hilbert space $\mathcal{K}_x$ and an isometry $V_x{:}\,\mathcal{H}_x(\breve{a}) \to \mathcal{K}_x$. Now the projections P_m in $\bigoplus_{m\in M_x} \mathcal{A}_{xm}(a) \otimes \mathcal{M}_m^x$ which correspond to the direct sum decomposition over m are mapped by π_x to projections on $\mathcal{K}_x$, so we get a decomposition into orthogonal

subspaces $\mathcal{K}_x = \bigoplus_m \pi_x(P_m)\,\mathcal{K}_x$. Since the P_m commute with all other elements of the algebra, the projections $\pi_x(P_m)$ commute with all $\pi_x(A)$, and $A \mapsto \pi_x(P_m)\pi_x(A)$ becomes a representation on $\pi_x(P_m)\,\mathcal{K}_x$. This representation can be restricted to the message algebra $\mathcal{M}_m^x$, and since the representation of a full matrix algebra is unique up to multiplicity (and up to unitary equivalence indicated by "$\simeq$" in the equations below), we can split the subspace $\pi_x(P_m)\,\mathcal{K}_x$ into a tensor product:

$$\begin{aligned} \pi_x(P_m)\,\mathcal{K}_x &\simeq \mathcal{H}_{xm}(\check{a}) \otimes \mathcal{K}_m^x\,, && (4.17)\\ \pi_x(\mathbb{1}\otimes X)\,\pi_x(P_m) &\simeq \mathbb{1}\otimes X\,, && (4.18)\\ \pi_x(A\otimes \mathbb{1})\,\pi_x(P_m) &\simeq \pi_{xm}(A)\otimes \mathbb{1}\,. && (4.19) \end{aligned}$$

At the last line we have used that all $\pi_x(A\otimes\mathbb{1})$ commute with all $\pi_x(\mathbb{1}\otimes X) \simeq (\mathbb{1}\otimes X)$, so must be of the form $A'\otimes\mathbb{1}$ for some $A' = \pi_{xm}(A)$. We have already indicated in the notation that the space $\mathcal{H}_{xm}(\check{a})$ arising in this construction will be chosen as Alice's lab Hilbert space for the coherent strategy $\check{a}$. The revert operation will simply be $R_{xm} = \pi_{xm}\colon \mathcal{A}_{xm} \to \mathcal{B}(\mathcal{H}_{xm}(\check{a}))$ and, finally, the isometries of the pure strategy will be

$$V_{x,m}(\check{a}) \simeq \pi_x(P_m)V_x(a)\colon \qquad \mathcal{H}_x(\check{a}) \to \pi_x(P_m)\,\mathcal{K}_x \simeq \mathcal{H}_{xm}(\check{a})\otimes\mathcal{K}_m^x\,. \tag{4.20}$$

Then Eq. (4.11) holds by virtue of the Stinespring representation, and we have shown that $\check{a}$ is indeed stronger than a.

To summarize: for every strategy a there is a stronger locally coherent strategy $\check{a}$. Moreover, the corresponding revert operation can be chosen to be a representation for all x. We will assume from now on that Bob uses coherent strategies, since this does not constrain his power to resolve Alice's actions at any stage.

4.3.3 Bounding Local Hilbert Space Dimensions

It is a crucial point in the definition of concealment that no limitations are imposed on Bob's capabilities. In particular, he could choose to use arbitrarily large local lab Hilbert spaces. In principle, this makes scanning all of Bob's strategies for checking ε-concealment an infinite task. However, the purification construction takes care of this aspect as well, and we will show that without loss of discrimination power Bob can fix the dimension of his lab spaces uniformly over all his strategies.

We have seen in Sec. 2.5 that the Stinespring construction respects finite-dimensionality. Choosing the "minimal" dilation, we have $\dim\mathcal{K} \leq \dim\mathcal{A}\cdot\dim\mathcal{H}$. However, since this bound still contains the algebra $\mathcal{A}$, which is part of the strategy whose purification generates the locally coherent protocol, and which is not a priori bounded, this argument does not suffice to derive a uniform dimension bound on local lab spaces.

The desired bound can be constructed by looking directly at the definition of locally

coherent strategies. Here the growth of Bob's lab space is given by the two operations

$$V_{x,m}(b) \quad : \quad \mathcal{H}_x(b) \to \mathcal{K}_m^x \otimes \mathcal{H}_{xm}(b) \tag{4.21}$$

at Bob's nodes and

$$\mathcal{H}_{xm}(b) \quad = \quad \mathcal{K}_m^x \otimes \mathcal{H}_x(b) \tag{4.22}$$

at Alice's nodes.

Given the dimensions of $\mathcal{H}_x(b)$ and $\mathcal{K}_m^x$, the first line per se does not imply a bound on the dimension of $\mathcal{H}_{xm}(b)$. However, the range of $V_{x,m}$ has known finite dimension, so most of these dimensions will never be used. More precisely, we can find a subspace $\mathcal{H}'_{xm}(b) \subset \mathcal{H}_{xm}(b)$ such that

$$V_{x,m}(b)\left(\mathcal{H}_x(b)\right) \subset \mathcal{K}_m^x \otimes \mathcal{H}'_{xm}(b). \tag{4.23}$$

Indeed, we can take $\mathcal{H}'_{xm}(b)$ as the span of all vectors $|\phi_{\alpha,j}\rangle$ appearing in the expansion $V_{x,m}(b)\,|\phi_\alpha\rangle = \sum_j |\psi_j\rangle \otimes |\phi_{\alpha,j}\rangle$, where $\{|\psi_j\rangle\} \subset \mathcal{K}_m^x$ and $\{|\phi_\alpha\rangle\} \subset \mathcal{H}_x(b)$ are orthonormal bases. Hence

$$\dim \mathcal{H}'_{xm}(b) \leq \dim \mathcal{H}_x(b)\, \dim \mathcal{K}_m^x\,. \tag{4.24}$$

We now apply this idea inductively, i.e., with a previously constructed $\mathcal{H}'_x(b) \subset \mathcal{H}_x(b)$ on the left hand side of Eq. (4.23). Note that at Alice's nodes there is nothing to choose, and the dimension bound Eq. (4.24) holds with equality anyhow. Moreover, at the root we have $\dim \mathcal{H}_0(b) = \dim \mathcal{H}'_0(b) = 1$ for all strategies.

Hence we have generated a new strategy, using the same isometries $V_{x,m}(b)$ as b, but with domains and ranges restricted to a subspace $\mathcal{H}_x(b') \equiv \mathcal{H}'_x(b) \subset \mathcal{H}_x(b)$ for all b. We will now show that b' is stronger than b. The required revert operation is implemented by the subspace embedding $j_x\colon \mathcal{H}_x(b') \to \mathcal{H}_x(b)$, as $R_x(B) = j_x^* B j_x$ and, due to Eq. (4.23), the operators $V_{x,m}$ for the new strategies are connected by

$$V_{xm}(b)\, j_x = (\mathbb{1} \otimes j_{xm})\, V_{xm}(b')\colon \quad \mathcal{H}_x(b') \to \mathcal{K}_m^x \otimes \mathcal{H}_{xm}(b)\,, \tag{4.25}$$

where j_{xm} is the embedding of $\mathcal{H}'_{xm}(b)$ into $\mathcal{H}_{xm}(b)$. Eq. (4.11) then follows by combining this with Eq. (4.15) in a version adapted to Bob's pure strategies. An intuitive description of this revert operation in the Schrödinger picture is to ask Bob to consider his density operator on $\mathcal{H}_x(b')$ as a density operator on the larger space $\mathcal{H}_x(b)$, by setting it equal to zero on the orthogonal complement.

It is perhaps paradoxical that in this case the strategy using less resources is stronger. But in fact, they are just equally strong. The revert operation in the opposite direction is $S_x : \mathcal{B}(\mathcal{H}_x(b')) \to \mathcal{B}(\mathcal{H}_x(b))$, with

$$S_x(B) = j_x B j_x^* + \rho_x(B)\,(\mathbb{1} - j_x j_x^*)\,, \tag{4.26}$$

where ρ_x is an arbitrary state on $\mathcal{B}(\mathcal{H}_x(b'))$. The second term is added to satisfy the channel normalization, $S_x(\mathbb{1}) = \mathbb{1}$. Since $j_x^* j_x = \mathbb{1}$, we have $R_x S_x = \mathrm{id}$. The revert operation in this case is thus the projection on the subspace $\mathcal{H}_x(b') \subset \mathcal{H}_x(b)$.

Taking together the reduction operation, and, possibly an expansion as described (adding some extra dimensions on which all states vanish), we can convert any strategy b to another one, for which the dimension bound Eq. (4.24) holds with equality, at both Bob's and Alice's nodes. But then we can identify all the spaces $\mathcal{H}_x(b')$ with a fixed space of appropriate dimension, say $\mathcal{H}_x^B$.

Applying the same construction to Alice's operations, we find a strategy-independent Hilbert space $\mathcal{H}_x^A$. In particular, we will henceforth assume $\mathcal{H}_x(\check{a}_0) = \mathcal{H}_x(\check{a}_1) = \mathcal{H}_x^A$ at all nodes x for Alice's locally coherent strategies $\check{a}_i$. This will simplify the discussion of Alice's cheating strategy in Secs. 4.3.5 and 4.3.7 below. We summarize these results in the following Proposition, which we formulate for Bob's strategies. It holds equally for Alice's strategies, too.

Proposition 4.4. (Dimension Bound)
Let $\mathcal{H}_x^B$ denote a family of Hilbert spaces with dimensions satisfying

$$\dim \mathcal{H}_{xm}^B = \dim \mathcal{H}_x^B \dim \mathcal{K}_m^x \tag{4.27}$$

$$\textit{and} \qquad \dim \mathcal{H}_0^B = 1 \tag{4.28}$$

for all nodes x. Then for every locally coherent strategy b of Bob there is an equally strong locally coherent strategy b' with $\mathcal{H}_x(b') = \mathcal{H}_x^B$ for all x.

The entire strategy dependence is now contained in the choice of the operators $V_{x,m}(b')$.

Corollary 4.5. *In the definition of ε-concealing and δ-cheating strategies in Sec. 4.2.2.3, we may restrict the quantifier over all of Bob's strategies to locally coherent strategies with a strategy-independent lab Hilbert space $\mathcal{H}_x^B$. For every $\xi > 0$ there is a finite set S of such strategies approximating all of Bob's discriminating procedures to within ξ. That is, for any strategy b of Bob we can find $b' \in S$ such that for all of Alice's strategies a:*

$$\|\rho_c(a,b) - \rho_c(a,b')\|_1 \leq \xi\,. \tag{4.29}$$

The proof of Cor. 4.5 is obvious from the dimension bound, and the observation that the set of bounded operators between Hilbert spaces of fixed finite dimension is compact in the norm topology.

4.3.4 Bob's Strategy Register

The next simplification we would like to introduce will significantly reduce the complexity of the many-round scenario. The basic idea is to replace all of Bob's choices by a single choice he makes at the beginning by preparing a suitable initial state. His later choices will then be taken over by a sequence of "quantum controlled operations". This reorganization of Bob's choices requires the expansion of the lab space by an additional

register, to hold the control information. It is perhaps worthwhile to emphasize that this strategy register serves merely as a technical tool in the no-go proof.

We will choose a finite approximation S to Bob's strategy space in the sense of Cor. 4.5, with a very small value of ξ, which will be taken to zero at the end. The strategy register will be described by the Hilbert space $\ell^2(S)$, the complex valued functions on S, with the usual scalar product. In other words, we have one basis vector $|b\rangle$ for each strategy $b \in S$. Then we set

$$\widetilde{\mathcal{H}}_x^B = \mathcal{H}_x^B \otimes \ell^2(S) \tag{4.30}$$

$$\widetilde{V}_{x,m} : \widetilde{\mathcal{H}}_x^B \to \widetilde{\mathcal{H}}_{xm}^B \otimes \mathcal{K}_m^x \tag{4.31}$$

$$\widetilde{V}_{x,m} = \sum_{b\in S} V_{x,m}(b) \otimes |b\rangle\langle b| \tag{4.32}$$

Observe that $\tilde{V}_{x,m}$ is now independent of Bob's strategy (it depends on S). However, Bob still has a choice to make, namely the choice of the initial state for the strategy register. If he wants to play strategy b, he will set it to $|b\rangle\langle b|$ and then let the pre-programmed controls take over.

The construction also opens up the rather interesting possibility for Bob to play strategies in superposition, simply by initially preparing a superposition of the basis states $|b\rangle$. For this case it is helpful to bear in mind that the "control" by "controlled unitary operations" is not a one way affair. As soon as Bob prepares superpositions, the strategy register is in general affected by the interaction, so by "measuring the strategy" after a while, Bob could pick up some clues about Alice's actions. This is required by basic laws of quantum mechanics, because the controlled-unitary operation creates entanglement.

Let us consider the overall effect of the protocol up to commitment, with Bob choosing an arbitrary initial state $\sigma \in \mathcal{B}_*(\ell^2(S))$ (possibly mixed) for the strategy register, and Alice playing strategy a. At commitment time, the total observable algebra is now $\bigoplus_{x\in X_c} \mathcal{A}_x(a) \otimes \mathcal{B}(\widetilde{\mathcal{H}}_x^B)$. The state obtained on this algebra depends linearly on the initial state σ, and being implemented by a series of completely positive transformations, this dependence is given by a quantum channel $\Gamma(a)$. In the Heisenberg picture we thus have

$$\Gamma(a) : \bigoplus_{x\in X_c} \mathcal{A}_x(a) \otimes \mathcal{B}(\widetilde{\mathcal{H}}_x^B) \to \mathcal{B}(\ell^2(S))\,. \tag{4.33}$$

The restriction of the final state to Bob's side is what decides his chances of distinguishing different strategies of Alice. These restrictions are given by the reduced channel $\Gamma^B(a) : \bigoplus_{x\in X_c} \mathcal{B}(\widetilde{\mathcal{H}}_x^B) \to \mathcal{B}(\ell^2(S))$, given by

$$\Gamma^B(a)\Big(\bigoplus_{x\in X_c} B_x\Big) = \Gamma(a)\Big(\bigoplus_{x\in X_c} \mathbb{1}_{\mathcal{A}_x(a)} \otimes B_x\Big)\,. \tag{4.34}$$

The concealment condition requires that $\Gamma^B(a_0) \approx \Gamma^B(a_1)$. The aim of the impossibility proof is to conclude from this the existence of a good cheating strategy for Alice. For

this conclusion it turns out to be crucial how the approximate equality of these channels is expressed quantitatively. We defer this discussion to Sec. 4.3.6, and treat first the case $\Gamma^B(a_0) = \Gamma^B(a_1)$, which requires only the Stinespring dilation theorem, and shows more clearly what properties we need to establish in the approximate case.

4.3.5 The Case of Perfect Concealment

In the sequel Bob is always understood to take advantage of his strategy register and pre-programmed controls, as described in Sec. 4.3.4. So we will henceforth drop the tilde on Bob's Hilbert spaces $\widetilde{\mathcal{H}}_x^B$ to streamline the presentation.

For the case of perfect concealment, suppose that $\Gamma^B(a_0) = \Gamma^B(a_1)$, and that Alice is preparing to cheat. She will then play the local purification $\check{a}_i$ $(i = 0, 1)$ of one of the honest strategies until commitment time. Note that both Alice's and Bob's strategies are assumed to be locally coherent in the sense of Sec. 4.3.2, with Hilbert space dimensions independent of their respective strategies as explained in Sec. 4.3.3. The concatenated channel $\Gamma(\check{a}_i)\colon \oplus_x \mathcal{B}(\mathcal{H}_x^A) \otimes \mathcal{B}(\mathcal{H}_x^B) \to \mathcal{B}(\ell^2(S))$ is then likewise pure, and is hence given by operators $V_{i,x} : \ell^2(S) \to \mathcal{H}_x^A \otimes \mathcal{H}_x^B$ as

$$\Gamma(\check{a}_i)\Big(\bigoplus_{x\in X_c}(A_x \otimes B_x)\Big) = \sum_{x\in X_c} V_{i,x}^*(A_x \otimes B_x)V_{i,x} = V_i^*\Big(\bigoplus_{x\in X_c}(A_x \otimes B_x)\Big)V_i\,. \quad (4.35)$$

In the last step of Eq. (4.35) we have combined all the $V_{i,x}$ into a single operator $V_i : \ell^2(S) \to \mathcal{K} := \bigoplus_x \mathcal{H}_x^A \otimes \mathcal{H}_x^B$, and the direct sum refers to the direct sum decomposition of the underlying Hilbert space $\mathcal{K}$. Note that this Hilbert space carries a representation π of Bob's observable algebra $\bigoplus_x \mathcal{B}(\mathcal{H}_x^B)$ at commitment time, simply by setting $\pi(\bigoplus_x B_x) = \bigoplus_x \mathbb{1}_x^A \otimes B_x$. Thus, $(\mathcal{K}, \pi, V_i)$ is a Stinespring dilation of the channel $\Gamma^B(\check{a}_i)$.

But now, by assumption $\Gamma^B(\check{a}_0) = \Gamma^B(a_0) = \Gamma^B(a_1) = \Gamma^B(\check{a}_1)$. Hence we get two dilations of the same channel, which must be connected by a unitary operator $U \in \mathcal{B}(\mathcal{K})$ as in Th. 2.7. Essentially, this U will be Alice's sneak flip operation. What we have to show is that she can execute this operation on the system under her control, given the classical information x.

The condition $U\pi(Y) = \pi(Y)U$, applied to a projection $Y = P_x$ of one of the summands implies that U can be broken into blocks, $U\pi(P_x) = \pi(P_x)U \in \mathcal{B}(\mathcal{H}_x^A \otimes \mathcal{H}_x^B)$. The intertwining relation for $\pi(B_x)$ allows us to conclude that this operator is of the form $U_x \otimes \mathbb{1}_x^B$, with a unitary operator $U_x \in \mathcal{B}(\mathcal{H}_x^A)$. Clearly, U_x is an operator between possible lab spaces of Alice, depending only on publicly available information $x \in X_c$. This will be Alice's cheat channel. Setting

$$C_x\colon \mathcal{B}(\mathcal{H}_x(\check{a}_1)) \to \mathcal{B}(\mathcal{H}_x(\check{a}_0)) \qquad C_x(A) = U_x^* A U_x\,, \quad (4.36)$$

we immediately conclude from $UV_0 = V_1$ that

$$\Gamma(\check{a}_0)(\bigoplus_x C_x \otimes \mathrm{id}_x^B) = \Gamma(\check{a}_1). \tag{4.37}$$

Let us summarize Alice's perfect cheat, as illustrated in Fig. 4.1. She will play the purification $\check{a}_0$ of the honest strategy a_0 until commitment time. If at that time she decides to go for the bit value 0, she will just apply the revert operation from the purification construction. After that nobody can tell the difference between her actions and the honest a_0, not even with full access to both labs. On the other hand, if she wants to choose bit value 1, she will apply the cheat channel C_x. We see from Eq. (4.37) that afterwards nobody will be able to tell the difference between her actions and $\check{a}_1$. Finally, she will apply the revert operation from $\check{a}_1$ to a_1, hiding all her tracks. Note that the revert operation by construction works at any step: indeed Alice can cheat at any time, since the protocol must be concealing for all steps in order to be concealing at the commitment stage.

4.3.6 Bob's Entangled Strategy Record

In the previous Section we have seen how Stinespring's theorem allows Alice to find a perfect cheat in a perfectly concealing bit commitment protocol. The continuity theorem presented in Sec. 4.3.7 below shows that the same cheating strategy still works for Alice with high probability under more realistic conditions — when only approximate concealment is guaranteed, $\Gamma^B(a_0) \approx \Gamma^B(a_1)$. The result crucially depends on the way in which the distance between these two channels is evaluated: Bob can test the condition $\Gamma^B(a_0) \approx \Gamma^B(a_1)$ by preparing a state σ for the strategy register $\ell^2(S)$, and making a measurement on the system $\mathcal{H}_x^B$ he receives back from Alice. This includes both the possibility to superpose his original strategies $|b\rangle$, and the possibility to mix such strategies in the sense of game theory. However, this still does not exhaust his options: he can *keep an entangled record of his strategy.* This would be pointless for just classical mixtures of his basic strategies $|b\rangle$. In that case all his density operators would commute with the "strategy observable", and he could extract the initial strategy by a von Neumann measurement from the state at any later step. However, if he also uses superpositions of strategies, the controlled unitaries may properly "change" the strategy. It therefore makes sense to keep a record, i.e., to not only use a mixed initial state, which would correspond to a mixed strategy in the sense of von Neumann's game theory, but to use an entangled pure state on $\ell^2(S) \otimes \ell^2(S')$, with some reference system S'. It turns out that one can always choose $S' \simeq S$ (cf. Prop. 8.11 in Paulsen's text [Pau02]). While the first copy in this tensor product is used as before to drive the conditional strategy operators V_x, the second is the record and is completely left out of the dynamics. In other words, Bob not only uses a von Neumann mixed strategy, but the purification of this mixture. Concealment will then have to be guaranteed against his joint measurements on $\mathcal{H}_B^x \otimes \ell^2(S')$.

We will see in Sec. 4.4.2 that this procedure in general does increase Bob's resolution for the difference of channels. Of course, if the initial selection of strategies S is large enough, an approximation of this quantum randomized strategy will already be contained in S, and the gain may be negligible. Mathematically, the introduction of randomized strategies corresponds to using a different norm: Alice will have to make sure that $\|\big(\Gamma^B(a_0)-\Gamma^B(a_1)\big)\otimes \mathrm{id}_n\|_\infty \leq \varepsilon$ if n-dimensional bystander systems are taken into account, for all $n \in \mathbb{N}$. As explained in Sec. 2.7.1, this just means that these two channels need to be indistinguishable in cb-norm, $\|\Gamma^B(a_0) - \Gamma^B(a_1)\|_{cb} \leq \varepsilon$ for some small $\varepsilon > 0$.

4.3.7 The Full Impossibility Proof

The full impossibility proof goes beyond the case of perfect concealment discussed in Sec. 4.3.5. It shows that Alice can still cheat if the bit commitment protocol is only approximately concealing, and provides explicit dimension-independent bounds on Alice's probability to pass Bob's tests undetected:

Theorem 4.6. (No-Go Theorem)
Any ε-concealing bit commitment protocol in the sense of Sec. 4.2.2 allows Alice to find a $2\sqrt{\varepsilon}$-cheating strategy.

These bounds coincide with those obtained by Spekkens and Rudolph [SR01] in the Kerckoffian framework. Our proof shows that they still hold if Bob no longer sticks to a publicly known strategy. This is a significant improvement over Cheung's dimension-dependent estimates [Che06], which do not suffice to rule out bit commitment protocols with large systems.

The full no-go proof is based on the continuity result for Stinespring's dilation, Th. 3.1. It states that any two quantum channels T_0, T_1 whose common domain and range are full matrix algebras are close in cb-norm iff there exist corresponding Stinespring isometries V_0, V_1 which are close in operator norm.

However, in our case the domain algebra of the commitment channels $\Gamma_i^B \equiv \Gamma^B(\check{a}_i)$ is not a full matrix algebra, but the direct sum $\oplus_x \mathcal{B}(\mathcal{H}_x^B)$. Again we have dropped the tilde from Bob's Hilbert spaces in an attempt to streamline the presentation. In order to apply the continuity theorem to our setting, we will extend the channels $\Gamma_i \equiv \Gamma(\check{a}_i)\colon \oplus_x \mathcal{B}(\mathcal{H}_x^A)\otimes\mathcal{B}(\mathcal{H}_x^B) \to \mathcal{B}(\mathcal{H})$ to channels $\hat{\Gamma}_0, \hat{\Gamma}_1\colon \mathcal{B}(\mathcal{H}^A\otimes\mathcal{H}^B) \to \mathcal{B}(\mathcal{K})$, where we have introduced the shortcuts $\mathcal{H} := \ell^2(S)$, $\mathcal{H}^A := \oplus_x \mathcal{H}_x^A$ and $\mathcal{H}^B := \oplus_x \mathcal{H}_x^B$. Note that the tensor product $\mathcal{H}^A \otimes \mathcal{H}^B$ has the direct sum decomposition $\oplus_{xy} \mathcal{H}_x^A \otimes \mathcal{H}_y^B$, and that $\oplus_x \mathcal{B}(\mathcal{H}_x^A \otimes \mathcal{H}_x^B)$ is the subalgebra in $\mathcal{B}(\mathcal{H}^A \otimes \mathcal{H}^B)$ which consists of those operators that are supported on the diagonal subspace $\oplus_x \mathcal{H}_x^A \otimes \mathcal{H}_x^B$. For direct sum channels $\Gamma_i(\oplus_x A_x \otimes B_x) = \sum_x V_{i,x}^*(A_x \otimes B_x)V_{i,x}$ as in Eq. (4.35), the extensions $\hat{\Gamma}_i = \hat{V}_i^*(\cdot)\hat{V}_i$

have Stinespring isometries $\hat{V}_0, \hat{V}_1 \colon \mathcal{H} \to \mathcal{H}^A \otimes \mathcal{H}^B = \oplus_{xy} \mathcal{H}^A_x \otimes \mathcal{H}^B_y$ given by

$$\hat{V}_i \psi := \bigoplus_{xy} \delta_{xy} V_{i,x} \psi \,. \tag{4.38}$$

The restrictions of $\hat{\Gamma}_i$ to Bob's output system $\mathcal{H}^B$ will be denoted by $\hat{\Gamma}^B_i$. We then have $\hat{\Gamma}^B_i = \Gamma^B_i \circ P$, where the completely positive map

$$P \colon \mathcal{B}(\mathcal{H}^B) \to \oplus_x \mathcal{B}(\mathcal{H}^B_x) \qquad P(B) = \oplus_x P_x B P_x \tag{4.39}$$

is composed of the projections P_x in $\mathcal{H}^B$ onto $\mathcal{H}^B_x$. Since

$$\|\hat{\Gamma}^B_0 - \hat{\Gamma}^B_1\|_{cb} = \|(\Gamma^B_0 - \Gamma^B_1) \circ P\|_{cb} \leq \|\Gamma^B_0 - \Gamma^B_1\|_{cb} \,, \tag{4.40}$$

we may now apply the left half of the continuity estimate Eq. (3.9) to the extended quantum channels $\hat{\Gamma}^B_i$ to conclude that

$$\inf_U \|(U \otimes \mathbb{1}_B)\hat{V}_0 - \hat{V}_1\|^2_\infty \leq \|\hat{\Gamma}^B_0 - \hat{\Gamma}^B_1\|_{cb} \leq \|\hat{\Gamma}_0 - \hat{\Gamma}_1\|_{cb} \,. \tag{4.41}$$

The minimization at this point is with respect to all unitary $U \in \mathcal{B}(\mathcal{H}^A)$, which can be given the block decomposition

$$U\psi = \bigoplus_x \sum_y U_{xy} \psi_y \tag{4.42}$$

with operators $U_{xy} : \mathcal{H}^A_y \to \mathcal{H}^A_x$. It turns out that the minimization in Eq. (4.41) can always be restricted to unitary operators whose off-diagonal blocks vanish. To see this, note that the left hand side of Eq. (4.41) can be rewritten as

$$\begin{aligned} \inf_U \|(U \otimes \mathbb{1}_B)\hat{V}_0 - \hat{V}_1\|^2_\infty &= \inf_U \sup_\varrho \operatorname{tr} \varrho \big(\hat{V}_0^*(U^* \otimes \mathbb{1}_B) - \hat{V}_1^*\big) \big((U \otimes \mathbb{1}_B)\hat{V}_0 - \hat{V}_1\big) \\ &= \inf_U \sup_\varrho \Big(2 - 2 \operatorname{Re} \operatorname{tr} \varrho\, \hat{V}_1^*(U \otimes \mathbb{1}_B)\hat{V}_0\Big) \,, \end{aligned} \tag{4.43}$$

where the supremum is taken over all states $\varrho \in \mathcal{B}_*(\mathcal{H})$. From the definition of the isometries $\hat{V}_i$ in Eq. (4.38) above it is straightforward to verify that

$$\hat{V}_1^*(U \otimes \mathbb{1}_B)\hat{V}_0 = \sum_x V^*_{1,x}(U_{xx} \otimes \mathbb{1}_x)V_{0,x} \tag{4.44}$$

in Eq. (4.43). Therefore, the minimization procedure on the left hand side of Eq. (4.41) is not affected by the off-diagonal blocks $\{U_{xy}, x \neq y\}$, which implies that the infimum is attained at a unitary operator that is a direct sum of unitaries, $U = \oplus_x U_x \in \oplus_x \mathcal{B}(\mathcal{H}^A_x)$. On the other hand, the cb-norm difference $\|\Gamma^B_0 - \Gamma^B_1\|_{cb}$ is easily seen to be upper bounded by $2\,\|(U \otimes \mathbb{1}_B)V_0 - V_1\|_\infty$ for any unitary operator $U = \oplus_x U_x$.

In summary, we have shown that the Continuity Theorem 3.1 can be extended to direct sum channels with a unitary U that respects the direct-sum decomposition:

Proposition 4.7. (Continuity Theorem for Direct Sum Channels) *Let $\mathcal{H}$ be a finite-dimensional Hilbert space, and let $\{\mathcal{H}_x^A\}_{x\in X}$ and $\{\mathcal{H}_x^B\}_{x\in X}$ be collections of finite-dimensional Hilbert spaces. Suppose that $V_0, V_1\colon \mathcal{H} \to \oplus_x \mathcal{H}_x^A \otimes \mathcal{H}_x^B$ are Stinespring isometries for the quantum channels $\Gamma_1, \Gamma_2\colon \oplus_x \mathcal{B}(\mathcal{H}_x^A \otimes \mathcal{H}_x^B) \to \mathcal{B}(\mathcal{H})$ such that*

$$\Gamma_i\Big(\bigoplus_x (A_x \otimes B_x)\Big) = \sum_x V_{i,x}^*(A_x \otimes B_x)V_{i,x} = V_i^*\Big(\bigoplus_x (A_x \otimes B_x)\Big)\, V_i\,. \tag{4.45}$$

Let $\Gamma_i^B\colon \oplus_x \mathcal{B}(\mathcal{H}_x^B) \to \mathcal{B}(\mathcal{H})$ be the local restrictions of the channels Γ_i to the system $\oplus_x \mathcal{B}(\mathcal{H}_x^B)$, given by $\Gamma_i^B(\oplus_x B_x) := V_i^*(\oplus_x \mathbb{1}_x^A \otimes B_x)V_i$. We then have:

$$\inf_U \|(U \otimes \mathbb{1}_B)V_0 - V_1\|_\infty^2 \leq \|\Gamma_0^B - \Gamma_1^B\|_{cb} \leq 2 \inf_U \|(U \otimes \mathbb{1}_B)V_0 - V_1\|_\infty\,, \tag{4.46}$$

where the minimization is over all unitary operators $U = \oplus_x U_x \in \oplus_x \mathcal{B}(\mathcal{H}_x^A)$.

The proof of the no-go theorem now immediately follows from Prop. 4.7.

Proof of Th. 4.6: Alice will play the purification $\check{a}_0$ of the honest strategy a_0 until commitment time. If at that time she decides to go for the bit value 0, she will just apply the revert operation R from the purification construction, as described in Sec 4.3.2. It is then no longer possible to tell the difference between her actions and the honest a_0, not even with full access to both labs. On the other hand, if she wants to choose bit value 1, she will apply the cheat channel $C_x\colon \mathcal{B}(\mathcal{H}_x(\check{a}_1)) \to \mathcal{B}(\mathcal{H}_x(\check{a}_0))$ given by $C_x(A) := U_x^* A U_x$, where $U := \oplus_x U_x \in \oplus_x \mathcal{B}(\mathcal{H}_x^A)$ is the unitary operator that attains the infimum in Eq. (4.46) above. Given an ε-concealing bit commitment protocol with local channels $\Gamma^B(a_i)$ satisfying $\|\Gamma^B(a_0) - \Gamma^B(a_1)\|_{cb} \leq \varepsilon$, the continuity estimate now implies that

$$\begin{aligned} \|\Gamma(\check{a}_0)\big(\bigoplus_x C_x \otimes \mathrm{id}_x^B\big) - \Gamma(\check{a}_1)\|_{cb} &\leq 2\,\|\left(U \otimes \mathbb{1}^B\right) V(\check{a}_0) - V(\check{a}_1)\|_\infty \\ &\leq 2\sqrt{\|\Gamma^B(a_0) - \Gamma^B(a_1)\|_{cb}} \\ &\leq\ 2\sqrt{\varepsilon}\,, \end{aligned} \tag{4.47}$$

where $V(\check{a}_0), V(\check{a}_1)$ are Stinespring isometries for $\Gamma(\check{a}_0)$ and $\Gamma(\check{a}_1)$, respectively. Since the cb-norm difference cannot increase under quantum channels, the same bound holds after Alice's revert operation R,

$$\|\Gamma(\check{a}_0)\big(\bigoplus_x C_x \otimes \mathrm{id}_x^B\big)\, R - \Gamma(a_1)\|_{cb} \leq 2\sqrt{\varepsilon}\,. \tag{4.48}$$

Alice can then confidently announce the bit value 1 in the opening. The probability of her cheat being detected is upper bounded by $2\sqrt{\varepsilon}$. This concludes the proof of the no-go theorem. ■

4.4 Protocols Relying on Decoherence

Here we describe provably secure bit commitment protocols relying on decoherence in Alice's lab (Sec. 4.4.1), Bob's lab (Sec. 4.4.2), or in the transmission line (Sec. 4.4.3).

4.4.1 The Trusted Coherence Shredder

We have already seen in Sec. 4.1 that a trusted third party makes perfect bit commitment a trivial task: Alice may submit the bit to an incorruptible notary public, who will store the bit in his vault throughout the holding phase, and later pass it on to Bob on Alice's notice. In this scenario, the notary public will have to be paid for the long-term safe storage of the bit. Clearly, Alice and Bob would get away with much lower fees if the notary's presence were only required once, and only as a witness, without even having to store a file about the event. Such a possibility is offered by quantum mechanics.

The basic idea is that the notary is present in Alice's lab until the end of the commitment phase, and sees to it that Alice plays honest. If the honest protocols were locally coherent, even that would be no help, since we have seen that Alice could carry out her cheating transformation later, in the holding phase. However, if the honest protocols (a_0, a_1) involve some measurement or other decoherence, the notary overseeing these actions can make a difference. He could prevent a later cheat by taking some part of the system with him and destroying it. In our example below it is even sufficient for him to just watch Alice make a measurement and, if he so chooses, to forget about the result straight away. The protocol is perfectly concealing, and is as binding as desired, if a dimension parameter d is chosen large enough.

The setting requires a d-dimensional Hilbert space, and two mutually unbiased orthonormal bases $\{|e_j\rangle\}_{j=1}^d$, $\{|f_k\rangle\}_{k=1}^d$, which means that $\langle e_j|e_k\rangle = \langle f_j|f_k\rangle = \delta_{jk}$, and $|\langle e_j|f_k\rangle|^2 = 1/d$, for all $j, k = 1, \ldots, d$. While the maximum number of mutually unbiased bases in a Hilbert space of given dimension d is the subject of ongoing research, here we only need two such bases, which are always easily constructed: starting from any given orthonormal basis $\{|e_j\rangle\}_{j=1}^d$, we may choose $\{|f_k\rangle\}_{k=1}^d$ as the Fourier-transformed basis,

$$|f_k\rangle := \frac{1}{\sqrt{d}} \sum_{j=1}^{d} e^{\frac{2\pi i}{d} jk} |e_j\rangle \,. \tag{4.49}$$

The protocol begins by Alice sending Bob half of the maximally entangled state

$$|\Omega\rangle = \frac{1}{\sqrt{d}} \sum_{j=1}^{d} |e_j\rangle \otimes |e_j\rangle = \frac{1}{\sqrt{d}} \sum_{j=1}^{d} |f_j\rangle \otimes |\overline{f_j}\rangle \,, \tag{4.50}$$

where $|\overline{f_j}\rangle$ denotes the complex conjugate of $|f_j\rangle$ with respect to the basis $\{|e_j\rangle\}_{j=1}^d$. Then, if she wants to commit the bit value "0", she makes a von Neumann measurement

in the basis $\{|e_j\rangle\}_{j=1}^d$, and records the result. Similarly, to commit a "1", she makes a measurement in the basis $\{|f_j\rangle\}_{j=1}^d$. Thus, if she plays honest, as vouched for by the notary public, she will have no quantum system left in her lab, only the classical information about the bit value, and her measurement result. This is the information she sends to Bob at the opening stage. To verify, he will make a measurement in the basis $\{|e_j\rangle\}_{j=1}^d$, if Alice claims to have submitted "0", and in the basis $\{|\overline{f_j}\rangle\}_{j=1}^d$ otherwise, finding the same result as Alice with probability 1.

The protocol is perfectly concealing, since in either case Bob gets a system in the chaotic state $\rho_B = \frac{1}{d}\mathbb{1}$. It is also binding, because whatever false bit value and measurement result Alice claims, Bob will confirm this only with probability $1/d$, i.e., practically never, if d is large.

This is essentially the bit commitment protocol originally proposed by Bennett and Brassard in 1984 [BB84]. Alice's EPR attack does not work in our scenario, since the notary public will not permit her to delay the measurements until after the commitment phase. There is also a variant of this protocol, in which the measurement is not actually carried out. In that case Alice prepares one of the mixed states

$$\rho_0 = \frac{1}{d}\sum_j |e_j \otimes e_j\rangle\langle e_j \otimes e_j| \,, \tag{4.51}$$

$$\rho_1 = \frac{1}{d}\sum_j |f_j \otimes \overline{f_j}\rangle\langle f_j \otimes \overline{f_j}| \,, \tag{4.52}$$

for committing "0" or "1", respectively. Now the notary watching her will see to it that she actually prepares these mixed states, and not their purifications. For verification Bob uses the support projections $P_{0,1} = d \cdot \rho_{0,1}$.

Once again, the protocol is perfectly concealing. Let us analyze Alice's cheating options, after she prepared ρ_0, with the trusted notary watching and then leaving. If she wants to change her commitment to "1", what she can do is to employ some local channel $T \otimes \mathrm{id}$ and hope to pass Bob's test with the projection P_1. The probability for this is

$$\begin{aligned}
\mathrm{tr}\,\rho_0(T\otimes \mathrm{id}\,)(P_1) &= \frac{1}{d}\sum_{k,j=1}^d \langle e_j, e_j|(T\otimes \mathrm{id}\,)(|f_k,\overline{f_k}\rangle\langle f_k,\overline{f_k}|)|e_j,e_j\rangle \\
&= \frac{1}{d}\sum_{k,j=1}^d |\langle e_j|\overline{f_k}\rangle|^2\ \langle e_j|T(|f_k\rangle\langle f_k|)|e_j\rangle \\
&= \frac{1}{d^2}\sum_{k,j=1}^d \langle e_j|T(|f_k\rangle\langle f_k|)|e_j\rangle \\
&= \frac{1}{d^2}\sum_{j=1}^d \langle e_j|T(\mathbb{1})|e_j\rangle \\
&= \frac{1}{d}\,.
\end{aligned} \tag{4.53}$$

The same computation applies to $\operatorname{tr} \rho_1(T \otimes \mathrm{id}\,)(P_0)$, so Alice's success probability is $1/d$ independent of her cheating channel, and may hence be chosen to be arbitrarily small by scaling up the dimension.

4.4.2 A Decoherence Monster in Bob's Lab

In the proof of Th. 4.6 we have shown that Alice has a cheating strategy for any concealing protocol. Hence it is not surprising that by weakening Alice's position, namely when decoherence eliminates her favorite cheating option, bit commitment protocols like those described in the previous Section become possible. But it may seem rather paradoxical that decoherence acting on Bob's side, presumably further hampering the weaker partner, can also lead to successful protocols.

Suppose that every morning, the cleaning service comes to Bob's lab, unplug all vacuum pumps, and restore what they take for tidiness. Only classical records survive this procedure. When Alice is convinced that she can rely on this happening, she might reassess her demands on concealment, and the two might agree on a bit commitment protocol, which under such circumstances is indeed both concealing and binding. This example shows very clearly that the entangled record introduced in the proof of the no-go theorem is essential.

The protocol we suggest relies on the distinction between the local erasure of information and the destruction of quantum correlations, as described in detail in Sec. 3.4. We emphasize again that this is a purely quantum-mechanical effect without an analogue in the classical world.

The protocol goes as follows: Bob initially supplies a pure state $|\psi\rangle$ on a d-dimensional Hilbert space $\mathcal{H}_B$. There is only one round for Alice, requiring her to send back a system with the same Hilbert space. Her honest strategies are specified by a pair of channels $T_k : \mathcal{B}(\mathcal{H}_A^k \otimes \mathcal{H}_B) \to \mathcal{B}(\mathcal{H}_B)$ $(k = 0, 1)$. We take them to be locally coherent, i.e., implemented by a single isometry $V_k : \mathcal{H}_B \to \mathcal{H}_A^k \otimes \mathcal{H}_B$ each. Their restrictions to Bob's side will be channels that witness a maximal separation between cb-norm and standard operator norm, as provided by Prop. 3.6: $T_0^B(X) = V_0^*(\mathbb{1} \otimes X)V_0 = R(X)$ is ε-randomizing, and $T_1^B(X) = V_1^*(\mathbb{1} \otimes X)V_1 = S(X)$ is completely depolarizing.

To reveal her commitment, Alice will later supply Bob with the ancilla system $\mathcal{H}_A^k$, alongside with the bit value k. Bob will then verify Alice's claim with a projective measurement on $V_k|\psi\rangle$, as illustrated in Fig. 4.3. Clearly, this protocol is perfectly *sound*, since Bob's measurement will confirm the bit value k with unit probability if both parties have followed their honest strategies. The protocol is $\frac{\varepsilon}{2}$-*concealing*, provided the decoherence monster strikes as planned, implementing some entanglement-breaking channel [HSR03] on Bob's reference system. By definition, these are the channels $D\colon \mathcal{B}(\mathcal{H}_B) \to \mathcal{B}(\mathcal{H}_B)$ such that $D_* \otimes \mathrm{id}\,(\varrho)$ is separable for any input state ϱ. Hence, these channels are sometimes also called *separable*. In Fig. 4.3, the decoherence

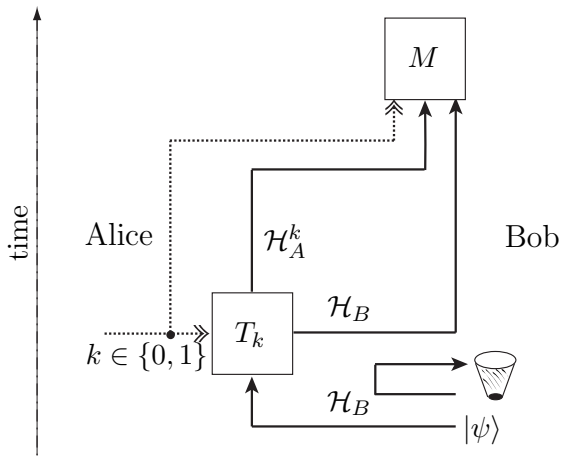

Figure 4.3: *A quantum bit commitment protocol with local decoherence in Bob's lab. The rubbish bin symbolizes an entanglement-breaking channel acting on Bob's reference system. The figure shows the flow of quantum (solid) and classical (dashed) information if both Alice and Bob play honest. Alice controls all systems on the left-hand side of the figure, Bob those on the right-hand side. Time flows upwards. The protocol starts with Bob submitting some pure quantum state $|\psi\rangle \in \mathbb{C}^d$ to Alice, and ends with Bob's measurement M.*

inflicted by D is indicated by the rubbish bin. We will show below that the maximal probability difference Bob can detect by preparing suitable states and making suitable measurements is then indeed just $\|R - S\|_\infty/2$.

To see that the protocol is binding, note first that Alice's usual cheating strategy cannot work: if there were an operator U such that $(U \otimes \mathbb{1})V_0 \approx V_1$ in norm, the two channels R and S could immediately be estimated to be cb-norm close, in contradiction to the second property guaranteed by Prop. 3.6.

However, it is clearly not enough to argue that there is no *universal* cheating strategy for Alice, which succeeds regardless of Bob's input state. We need to rule out strategies which would allow Alice to fool Bob's test in many cases, or with high probability. In addition, we also have to show security for arbitrary cheating strategies and, in particular, we have to make certain that the reduction of Bob's lab capabilities by the decoherence monster does not also give Alice a bit more freedom to cheat. That is, in order to prove security we have to explain why the coherent record makes a difference for Bob's ability to distinguish the honest strategies, but not for his ability to distinguish honest from cheating strategies in the opening phase. This is the essence of the following

Theorem 4.8. *Let $\varepsilon > 0$, $\delta > 0$. Then for sufficiently large dimension d the bit commitment protocol described above is perfectly sound, ε-concealing, and δ-binding.*

Let us focus on the concealment part first. In the protocol we grant the decoherence monster the freedom to apply an arbitrary *entanglement-breaking* quantum channel on Bob's bystander system. Any such channel $D\colon \mathcal{B}(\mathcal{H}_B) \to \mathcal{B}(\mathcal{H}_B)$ can be decomposed [HSR03] as $D = D_1 \circ D_2$, where

$$D_1\colon \mathcal{C}_X \to \mathcal{B}(\mathcal{H}_B) \qquad \text{and} \tag{4.54}$$

$$D_2\colon \mathcal{B}(\mathcal{H}_B) \to \mathcal{C}_X, \tag{4.55}$$

for some Abelian algebra $\mathcal{C}_X$. In other words, any entanglement-breaking channel can be thought of as being built from a measurement channel D_1, with resulting classical output system $\mathcal{C}_X$, followed by a re-preparation D_2.

In order to confirm ε-concealment of the monster protocol, we will need to show that any such entanglement-breaking channel D renders Bob's bystander system useless for the analysis of Alice's actions:

Lemma 4.9. *For any linear map $L\colon \mathcal{B}(\mathcal{H}) \to \mathcal{B}(\mathcal{K})$ and any entanglement-breaking channel $D\colon \mathcal{B}(\mathcal{H}_1) \to \mathcal{B}(\mathcal{K}_1)$,*

$$\|L \otimes D\|_\infty = \|L\|_\infty \,. \tag{4.56}$$

Since entanglement-breaking channels have a decomposition $D_1 \circ D_2$ with an intermediate classical system $\mathcal{C}_X$, it will turn out to be sufficient to verify this property for the noiseless classical channel id_X:

Lemma 4.10. *For any linear map $L\colon \mathcal{B}(\mathcal{H}) \to \mathcal{B}(\mathcal{K})$ and any classical observable algebra $\mathcal{C}_X$,*

$$\|L \otimes \mathrm{id}_X\|_\infty = \|L\|_\infty \,. \tag{4.57}$$

Proof of Lemma 4.10: For any $A \in \mathcal{B}(\mathcal{H})$ we have,

$$\|L(A)\|_\infty = \|L \otimes \mathrm{id}_X\,(A \otimes \mathbb{1}_X)\|_\infty \leq \|L \otimes \mathrm{id}_X\|_\infty \, \|A\|_\infty \,, \tag{4.58}$$

which shows that $\|L\|_\infty \leq \|L \otimes \mathrm{id}_X\|_\infty$.

For the converse implication, recall from Prop. 2.2 that any classical-quantum state ϱ on $\mathcal{B}(\mathcal{K}) \otimes \mathcal{C}_X$ can be given the form

$$\varrho = \sum_{x=1}^{|X|} p_x \, \varrho_x \otimes |x\rangle\langle x| \,, \tag{4.59}$$

where $\{p_x\}_{x=1}^{|X|}$ is a classical probability distribution, $\{\varrho_x\}_{x=1}^{|X|}$ is a set of quantum states

on $\mathcal{B}(\mathcal{K})$, and $\{|x\rangle\}_{x=1}^{|X|}$ denotes an orthonormal basis for $\mathbb{C}^{|X|}$. We may now estimate,

$$\begin{aligned}\|(L_* \otimes \mathrm{id}_X)\varrho\|_1 &\leq \sum_{x=1}^{|X|} p_x \, \|L_*(\varrho_x) \otimes |x\rangle\langle x|\|_1 \\ &= \sum_{x=1}^{|X|} p_x \, \|L_*(\varrho_x)\|_1 \\ &\leq \sum_{x=1}^{|X|} p_x \, \|L\|_\infty = \|L\|_\infty \,,\end{aligned} \tag{4.60}$$

and hence $\|L \otimes \mathrm{id}_X\|_\infty \leq \|L\|_\infty$, as claimed. ■

Proof of Lemma 4.9: Choosing $A \in \mathcal{B}(\mathcal{H})$, we immediately have

$$\begin{aligned}\|L(A)\|_\infty &= \|L(A) \otimes \mathbb{1}_{\mathcal{K}_1}\|_\infty \\ &= \|(L \otimes D)\,(A \otimes \mathbb{1}_{\mathcal{H}_1})\|_\infty \\ &\leq \|L \otimes D\|_\infty \, \|A \otimes \mathbb{1}_{\mathcal{H}_1}\|_\infty \\ &= \|L \otimes D\|_\infty \, \|A\|_\infty \,,\end{aligned} \tag{4.61}$$

implying $\|L\|_\infty \leq \|L \otimes D\|_\infty$.

For the converse implication, let $D = D_1 \circ D_2$ be a decomposition as in Eqs. (4.54) and (4.55) above. We may then estimate,

$$\begin{aligned}\|L \otimes D\|_\infty &= \|L \otimes (D_1 \circ D_2)\|_\infty \\ &= \|(\mathrm{id}_{\mathcal{K}} \otimes D_1)\,(L \otimes \mathrm{id}_X)\,(\mathrm{id}_{\mathcal{H}} \otimes D_2)\|_\infty \\ &\leq \|\mathrm{id}_{\mathcal{K}} \otimes D_1\|_\infty \, \|L \otimes \mathrm{id}_X\|_\infty \, \|\mathrm{id}_{\mathcal{H}} \otimes D_2\|_\infty \\ &\leq \|D_1\|_{cb} \, \|L \otimes \mathrm{id}_X\|_\infty \, \|D_2\|_{cb} = \|L\|_\infty \,,\end{aligned} \tag{4.62}$$

where in the last step we have used Lemma 4.10 and the fact that $\|T\|_{cb} = 1$ for any channel T (cf. Sec. 2.7.1). ■

We now have all the tools at hand to complete the

Proof of Th. 4.8: Soundness of the protocol is clear. Setting $L := R - S$ in Lemma 4.9, ε-concealment follows immediately from Prop. 3.6.

Thus, it only remains to show that the protocol is δ-binding. As a warm-up exercise, let us first exclude the possibility of Alice committing to the bit value k in the commitment phase, and then announcing the bit $1-k$ in the opening phase. This is sometimes called *passive cheating*.

If Bob has initially supplied the pure state $|\psi\rangle \in \mathbb{C}^d$, the probability of successfully passing Bob's projective measurement in such a scenario is $P(\psi) := |\langle V_0\psi|V_1\psi\rangle|^2$, resulting in the overall cheating probability

$$P := \int P(\psi)\, d\psi = \int \langle\psi|V_0^* V_1(|\psi\rangle\langle\psi|)V_1^* V_0|\psi\rangle \, d\psi \overset{(2.45)}{=} \overline{F}(V_0^* V_1)\,. \tag{4.63}$$

For δ and d as in Prop. 3.6, we then have the estimate

$$\begin{aligned}
2-\delta &\overset{(3.47)}{\leq} \|T_{0*}^B \otimes \mathrm{id}\,(|\Omega\rangle\langle\Omega|) - T_{1*}^B \otimes \mathrm{id}\,(|\Omega\rangle\langle\Omega|)\|_1 \\
&\leq \|(V_0\otimes\mathbb{1})|\Omega\rangle\langle\Omega|(V_0^*\otimes\mathbb{1}) - (V_1\otimes\mathbb{1})|\Omega\rangle\langle\Omega|(V_1^*\otimes\mathbb{1})\|_1 \\
&\overset{(2.48)}{\leq} 2\sqrt{1-f^2(V_0\otimes\mathbb{1}|\Omega\rangle, V_1\otimes\mathbb{1}|\Omega\rangle)} \\
&\overset{(2.44)}{=} 2\sqrt{1-F_c(V_0^*V_1)} \\
&\overset{(2.46)}{\leq} 2\sqrt{1-\overline{F}(V_0^*V_1)+\frac{1}{d}} \\
&\overset{(4.63)}{=} 2\sqrt{1-P+\frac{1}{d}}\,, \qquad (4.64)
\end{aligned}$$

where in the second step we have used that the trace-norm cannot increase under the partial trace operation [NC00]. From Eq. (4.64) we conclude that

$$P \leq \frac{1}{d} + \delta\,. \qquad (4.65)$$

Since the right side of Eq. (4.65) can be made as small as desired by stepping up the dimension, this gives the desired upper bound on Alice's probability of successfully passing Bob's test.

So far we have only proven bindingness against passive cheating attacks. As illustrated in Fig. 4.4, Alice's most general attack consists of applying some quantum channel $T^\sharp\colon \mathcal{B}(\mathcal{H}_\sharp)\otimes\mathcal{B}(\mathcal{H}_B)\to\mathcal{B}(\mathcal{H}_B)$ during the commitment phase, independently of the bit value $k\in\{0,1\}$. She will send a d-dimensional quantum system $\mathcal{H}_B$ to Bob without having committed to either bit. Only before the opening, she will then decide on a bit value k, apply a corresponding quantum channel $T_k^\sharp\colon \mathcal{B}(\mathcal{H}_A^k)\to\mathcal{B}(\mathcal{H}_\sharp)$ on her remaining system, and hope to pass Bob's projective measurement.

Assuming that Alice is a not prejudiced towards either bit, the probability of passing Bob's test is then $P := \frac{1}{2}P_0 + \frac{1}{2}P_1$, where for $k\in\{0,1\}$ we set

$$P_k := \int \langle V_k\psi|(T_{k*}^\sharp\otimes\mathrm{id}_B)T_*^\sharp(|\psi\rangle\langle\psi|)|V_k\psi\rangle\; d\psi. \qquad (4.66)$$

This probability can be bounded as follows:

$$\begin{aligned}
P_k &= \int \langle\psi|V_k^*(T_{k*}^\sharp\otimes\mathrm{id}_B)T_*^\sharp(|\psi\rangle\langle\psi|)V_k|\psi\rangle\; d\psi \\
&\overset{(2.45)}{=} \overline{F}\big(V_k^*(T_{k*}^\sharp\otimes\mathrm{id}_B)T_*^\sharp V_k\big) \\
&\overset{(2.46)}{\leq} F_c\big(V_k^*(T_{k*}^\sharp\otimes\mathrm{id}_B)T_*^\sharp V_k\big) + \frac{1}{d} \\
&\overset{(2.44)}{=} f^2\big(V_k\otimes\mathbb{1}_{B'}|\Omega\rangle,\, (T_{k*}^\sharp\otimes\mathrm{id}_B\otimes\mathrm{id}_{B'})(T_*^\sharp\otimes\mathrm{id}_{B'})(|\Omega\rangle\langle\Omega|)\big) + \frac{1}{d} \\
&\leq f^2\big(T_{k*}^B\otimes\mathrm{id}_{B'}(|\Omega\rangle\langle\Omega|), \mathrm{tr}_{\mathcal{H}_\sharp}T_*^\sharp\otimes\mathrm{id}_{B'}(|\Omega\rangle\langle\Omega|)\big) + \frac{1}{d}\,, \qquad (4.67)
\end{aligned}$$

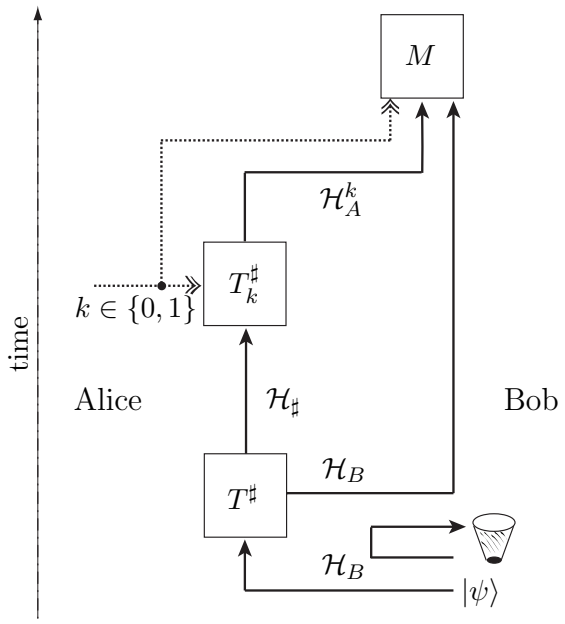

Figure 4.4: *Alice's cheating strategy consists of applying some quantum channel $T^\sharp$ in the commitment phase, and then another quantum channel $T_k^\sharp$ to commit to the bit value $k \in \{0,1\}$ only before the opening. Her goal is to pass Bob's projective measurement M.*

where in the final step we have used the monotonicity of the fidelity under the partial trace operation. Combining this estimate with Prop. 2.12 and Eq. (3.47) then immediately yields the bound

$$P \leq \frac{1}{2} + \frac{1}{d} + \frac{1}{2} f\big(R \otimes \mathrm{id}_{B'}(|\Omega\rangle\langle\Omega|), S \otimes \mathrm{id}_{B'}(|\Omega\rangle\langle\Omega|)\big) \leq \frac{1}{2} + \frac{1}{d} + \frac{1}{2}\sqrt{\delta}\,. \tag{4.68}$$

The RHS of Eq. (4.68) can be brought as close to $\frac{1}{2}$ as desired by stepping up the dimension. Resubstituting $\frac{1}{d} + \frac{1}{2}\sqrt{\delta} \mapsto \delta$, the protocol is δ-binding. This concludes the proof of Th. 4.8. ■

4.4.3 Decoherence in the Transmission Line

While noise in the transmission line is generally considered a nuisance, and coding theorists have designed elaborate error correcting codes to cope with it, Wyner [Wyn75] was the first to realize that noise may sometimes be beneficial for cryptographic applications — in his case for key distribution. Crépeau and Kilian [CK88] have later shown that classical noisy channels may also be employed to establish secure bit commitment. Their results have subsequently been improved in [Cre97, DKS99]. Recently Winter *et al.* [WNI03] have considered the asymptotic version of string commitment and have obtained a single-letter expression for the commitment capacity of a classical noisy

channel. Their results show that any nontrivial noisy channel can be used to establish secure bit commitment. The theorem can be extended to so-called *classical-quantum* channels. But it remains an open question whether fully quantum channels can also be useful for bit commitment.

Misaligned spatial reference frames can also effectively act as a noisy channel, and facilitate secure bit commitment. An example for a secure protocol was recently given by Harrow *et al.* [HOT06].

4.5 Quantum Bit Commitment with Continuous-Variable Systems

We have so far restricted the discussion of the no-go theorem to systems that can be described in finite-dimensional (albeit arbitrarily large) Hilbert spaces. In this Section we show that the results can be easily extended to continuous variable systems — as long as the systems obey a global energy constraint of a reasonably generic form. The total available energy for the protocol needs to be finite but can otherwise be as high as desired, and yet quantum bit commitment remains impossible. Purists might dismiss this additional energy constraint on the basis that it restricts the domain for the impossibility proof. Yet most physicists know that infinite energy is seldom available. We do not yet know if the no-go theorem applies to continuous variable systems with unbounded energy also.

To set the stage, assume that $\mathcal{H}$ is a separable (but no longer necessarily finite-dimensional) Hilbert space. As before, let $\mathcal{B}_*(\mathcal{H})$ denote the Banach space of trace-class operators on $\mathcal{H}$, and $\mathcal{S}(\mathcal{H}) \subset \mathcal{B}_*(\mathcal{H})$ the closed convex set of states. We further assume that $H\colon \mathcal{D} \to \mathcal{H}$ is an unbounded self-adjoint (energy) operator defined on a dense set $\mathcal{D} \subset \mathcal{H}$. (From the Hellinger-Toeplitz theorem (cf. Sec. III.4 in [RS80]) we know that a symmetric unbounded operator cannot be defined on all of $\mathcal{H}$, so we always assume a dense subset $\mathcal{D}$.) For the proof we assume that H has discrete spectrum, that all of its eigenvalues h_n have finite multiplicity, and that $\lim_{n\to\infty} h_n = \infty$. Under these conditions, the set of states

$$\mathcal{S}_E(\mathcal{H}) := \big\{ \varrho \in \mathcal{S}(\mathcal{H}) \;\big|\; \operatorname{tr} \varrho\, H \leq E \big\} \tag{4.69}$$

can be shown to be compact for every $E \geq 0$ [Hol03]. As we assume this energy constraint to be global, we impose that it is respected by the quantum operation T_* that describes the full bit commitment protocol: $T_*(\varrho) \in \mathcal{S}_E(\mathcal{H})$ for all $\varrho \in \mathcal{S}_E(\mathcal{H})$.

As explained in Sec. 3.6, the continuity theorem applies in this setting. The proof presented in Sec. 4.3 then goes trough unchanged. There is also a simpler proof, which avoids the compactness arguments and is based on a useful approximation result: any infinite-dimensional system with energy constraints as in Eq. (4.69) can be approximated to arbitrary degree of accuracy by a sufficiently large finite-dimensional system.

This allows to reduce any bit commitment protocol to its finite-dimensional counterpart:

Proposition 4.11. (Reduction)
Given an ε-concealing and δ-binding quantum bit commitment protocol with a global energy constraint as in Eq. (4.69). Then for any $\gamma > 0$ there is a corresponding protocol on finite-dimensional Hilbert spaces with some dimension $d = d(\gamma)$ which is $(\varepsilon + \gamma)$-concealing and $(\delta + \gamma)$-binding.

Since the latter protocol is unfeasible for sufficiently small parameters ε, δ and γ, so is the former. The finite-dimensional approximation needed for the proof of Prop. 4.11 relies on the following two lemmas:

Lemma 4.12. *Let $\gamma > 0$ and $\mathcal{S}_E(\mathcal{H})$ as in Eq. (4.69). Then there exists a finite-dimensional projector P_γ such that*

$$\operatorname{tr} \varrho P_\gamma \geq 1 - \gamma \qquad \forall \quad \varrho \in \mathcal{S}_E(\mathcal{H}). \tag{4.70}$$

As a consequence, every system with energy constraints is essentially supported on a finite-dimensional Hilbert space.

Lemma 4.13. *Let $\gamma > 0$ and P_γ as in Lemma 4.12. Then for every quantum channel $T_*: \mathcal{B}_*(\mathcal{H}) \to \mathcal{B}_*(\mathcal{H})$ which respects the energy constraint Eq. (4.69) we have*

$$\|T_*(\varrho) - \frac{1}{\operatorname{tr} P_\gamma T_*(P_\gamma \varrho P_\gamma)} P_\gamma \, T_*(P_\gamma \varrho P_\gamma) \, P_\gamma\|_1 \leq 4\sqrt{\gamma} + \frac{2\gamma}{1-\gamma} \tag{4.71}$$

for all $\varrho \in \mathcal{S}_E(\mathcal{H})$.

The proof of Prop. 4.11 is then straightforward: Given the continuous-variable bit commitment protocol with energy bound E and security parameters ε and δ, we construct its finite-dimensional companion by projecting on the subspace $P_\gamma \mathcal{H}$, with the finite-dimensional projector P_γ chosen as in Lemma 4.12. We know from the discussion in Sec. 4.3 that both the concealment and the bindingness condition can be expressed in terms of appropriately chosen quantum channels T_*. By assumption, these will respect the energy constraint. The approximation in Lemma 4.13 then guarantees that for sufficiently small γ the companion protocol has nearly identical security parameters. Substituting $4\sqrt{\gamma} + \frac{2\gamma}{1-\gamma} \mapsto \gamma$, this concludes the proof.

It remains to derive the approximation lemmas. The proof of Lemma 4.12 appears in [Hol03]. We include it here for completeness:

Proof of Lemma 4.12: Let the eigenvalues of the energy operator H be arranged in increasing order: $h_1 \leq h_2 \leq h_3 \leq \ldots$, with eigenprojector P_n corresponding to the

eigenvalue h_n. For $N \in \mathbb{N}$, we set $\hat{P}_N := \sum_{n=1}^{N} P_n$. We then have for all $\psi \in \mathcal{H}$:

$$\begin{aligned} \langle\psi|h_{N+1}\big(\mathbb{1} - \hat{P}_N\big)|\psi\rangle &= \langle\psi|h_{N+1} \sum_{n=N+1}^{\infty} P_n|\psi\rangle \\ &\leq \langle\psi| \sum_{n=N+1}^{\infty} h_n P_n|\psi\rangle \\ &\leq \langle\psi|H|\psi\rangle, \end{aligned} \tag{4.72}$$

implying that $h_{N+1}(\mathbb{1} - \hat{P}_N) \leq H$ for all $N \in \mathbb{N}$. We may then conclude that

$$\operatorname{tr} \varrho\big(\mathbb{1} - \hat{P}_N\big) \leq \frac{1}{h_{N+1}} \operatorname{tr} \varrho\, H \leq \frac{E}{h_{N+1}} \tag{4.73}$$

for all $\varrho \in \mathcal{S}_E(\mathcal{H})$. Since the sequence $\{h_N\}_{N\in\mathbb{N}}$ diverges, the result follows by choosing $P_\gamma := \hat{P}_{N_0}$ for some sufficiently large $N_0 \in \mathbb{N}$. ■

Proof of Lemma 4.13: An application of the triangle inequality shows that

$$\begin{aligned} \|\varrho - P_\gamma \varrho P_\gamma\|_1 &\leq \|\varrho - P_\gamma \varrho\|_1 + \|P_\gamma \varrho - P_\gamma \varrho P_\gamma\|_1 \\ &\leq \|(\mathbb{1} - P_\gamma)\varrho\|_1 + \|\varrho(\mathbb{1} - P_\gamma)\|_1 \,. \end{aligned} \tag{4.74}$$

For $\varrho \in \mathcal{S}_E(\mathcal{H})$ we know from Lemma 4.12 that $\operatorname{tr}(\mathbb{1} - P_\gamma)\varrho \leq \gamma$, and thus the two terms on the RHS of Eq. (4.74) may be bounded as follows:

$$\begin{aligned} \|\big(\mathbb{1} - P_\gamma\big)\varrho\|_1 &= \operatorname{tr} U\big(\mathbb{1} - P_\gamma\big)\varrho \\ &\leq \operatorname{tr}^{\frac{1}{2}} \sqrt{\varrho} U U^* \sqrt{\varrho}\ \operatorname{tr}^{\frac{1}{2}} \sqrt{\varrho}\big(\mathbb{1} - P_\gamma\big)\sqrt{\varrho} \\ &\leq \sqrt{\gamma}\,, \end{aligned} \tag{4.75}$$

where we have used the Cauchy-Schwarz inequality for the Hilbert-Schmidt inner product, and U denotes the polar isometry of the operator $(\mathbb{1} - P_\gamma)\varrho$. Analogously, we have $\|\varrho(\mathbb{1} - P_\gamma)\|_1 \leq \sqrt{\gamma}$, which together with Eqs. (4.74) and (4.75) implies that

$$\|\varrho - P_\gamma \varrho P_\gamma\|_1 \leq 2\sqrt{\gamma}\,. \tag{4.76}$$

For all $\varrho \in \mathcal{S}_E(\mathcal{H})$, the renormalized state $\frac{1}{\operatorname{tr} P_\gamma \varrho} P_\gamma \varrho P_\gamma$ satisfies the estimate

$$\frac{1}{\operatorname{tr} P_\gamma \varrho} P_\gamma \varrho P_\gamma - P_\gamma \varrho P_\gamma \leq \frac{\gamma}{1-\gamma} P_\gamma \varrho P_\gamma\,, \tag{4.77}$$

which in combination with Eq. (4.76) implies that

$$\|\varrho - \frac{1}{\operatorname{tr} P_\gamma \varrho} P_\gamma \varrho P_\gamma\|_1 \leq 2\sqrt{\gamma} + \frac{\gamma}{1-\gamma}. \tag{4.78}$$

Since the trace norm cannot increase under quantum operations [NC00], the upper bound also holds for the norm difference $\|T_*(\varrho) - \frac{1}{\operatorname{tr} P_\gamma \varrho} T_*(P_\gamma \varrho P_\gamma)\|_1$. As the quantum channel T is supposed to respect the energy constraint Eq. (4.69), an analogous chain of estimates for the output states of the channel and yet another application of the triangle inequality then yield the desired result. ■

Chapter 5

Quantum Channel Capacities

Quantum channel capacity is one of the key quantitative notions of quantum information theory. Whenever one asks "how much quantum information" can be stored in a device, or sent down a transmission line, the ultimate answer is given by the quantum channel capacity: it is the maximal number of qubit transmissions per use of the channel, taken in the limit of long messages and using error correction schemes asymptotically eliminating all transmission errors.

In this Chapter we present a brief overview of the theory of quantum channel capacities. We mostly focus on the capacity for quantum information, but also address classical information transfer over quantum channels and additional resources, like classical side channels and entanglement. We restrict the discussion to *memoryless* channels, in which successive channel inputs are acted on independently. Memory effects will then be investigated in detail in Ch. 6. The interrelations between quantum channel capacity and distillable entanglement are studied in Sec. 8.7.

The presentation in this Chapter is based on the review article [Kre06]. Some of the results have appeared in a joint paper with R. F. Werner [KW04].

5.1 Introduction and Overview

We know from Sec. 2.4 that any processing of quantum information, be it storage or transfer, can be represented as a quantum channel: a completely positive and trace-preserving map that transforms states (density matrices) on the sender's end of the channel into states on the receiver's end. Very often the channel S that sender and receiver (conventionally called Alice and Bob, respectively) would like to implement is not readily available, typically due to detrimental noise effects, limited technology, or insufficient funding. They may then try to simulate S with some other channel T, which they happen to have at their disposal. The quantum channel capacity $Q(T, S)$

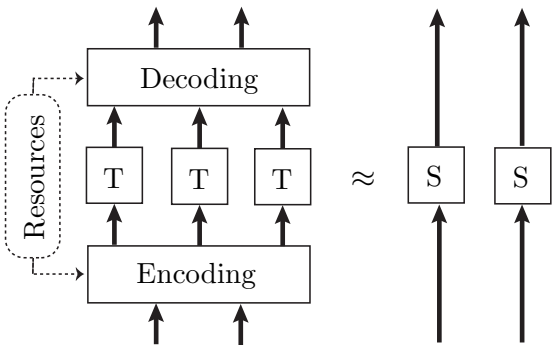

Figure 5.1: *Equipped with collective encoding and decoding operations (and perhaps some auxiliary resources), $n = 3$ instances of the channel T simulate $m = 2$ instances of the channel S. The transmission rate of the above scheme is $2/3$. Capacity is the largest such rate, in the limit of long messages and optimal encoding and decoding.*

of T with respect to S quantifies how well this simulation can be performed, in the limit of long input strings, so that Alice and Bob can take advantage of collective pre- and post-processing (cf. Fig. 5.1). Higher capacities may result if Alice and Bob are allowed to use additional resources in the process, such as classical side channels or a bunch of maximally entangled pairs shared between them.

Quantum capacity thus gives the ultimate benchmarks for the simulation of one quantum channel by another and for the optimal use of auxiliary resources. Together with the compression rate of a quantum source (cf. Secs. 2.8 and 7.3), it lies at the heart of quantum information theory.

In a very typical scenario Alice and Bob would like to implement the ideal (noiseless) quantum channel $S = \mathrm{id}$: they are interested in sending quantum states undistorted over some distance, or store them safely for some period of time, so that all the precious quantum correlations are preserved. The capacity $Q(T) \equiv Q(T, \mathrm{id})$ is then the maximal number of qubit transmissions per use of the channel, taken in the limit of long messages and using collective encoding and decoding schemes asymptotically eliminating all transmission errors. This is what is generally called the *quantum capacity* of the channel T, and is our main focus in this Chapter. Not too much is known so far about the quantum capacity for the simulation of other (non-ideal) channels (cf. Sec. 5.4).

In remarkable contrast to the classical setting, quantum channel capacities are very much affected by additional resources. This leads to unexpected and fascinating applications such as teleportation [BBC$^+$93] and superdense coding [BW92]. But it also results in a bewildering variety of inequivalent channel capacities, which still hold many challenges for future research.

This Chapter is organized as follows: In Sec. 5.2 we start with a rigorous definition of quantum channel capacity and explain some of its variations. Sec. 5.3 then explores

some elementary but useful properties of the quantum channel capacity. In Sec. 5.4 we discuss capacities for classical information transfer and capacities enhanced by additional resources. Sec. 5.5 shows how the various channel capacities can be expressed in terms of entropic information measures, and also includes a sketch of Devetak's proof [Dev05] of the quantum channel coding theorem in Sec. 5.5.4.

Our discussion in this Chapter is restricted to *memoryless* quantum channels, which are characterized by the requirement that successive channel inputs are acted on independently: messages of n symbols are processed by the tensor product channel $T^{\otimes n}$ (cf. Fig. 5.1), and hence there are no correlations between consecutive channel uses. This assumption considerably simplifies the analysis, but is often not justified in real-world applications. Memory effects are investigated in detail in Ch. 6.

In this Chapter and throughout this thesis we mainly concentrate on channels between finite-dimensional systems. This is enough to bring out the basic ideas. Many of the concepts and results discussed here can be generalized to *Gaussian channels*, which play a central role as building blocks for quantum optical communication lines [HW01, EW05].

5.2 Capacity for Quantum Information

The intuitive concept underlying quantum channel capacity is made rigorous in the following

Definition 5.1. (Quantum Channel Capacity)
A number $r \geq 0$ is called achievable rate *for the quantum channel $T\colon \mathcal{B}(\mathcal{H}_B) \to \mathcal{B}(\mathcal{H}_A)$ with respect to the quantum channel $S\colon \mathcal{B}(\mathcal{H}_{B'}) \to \mathcal{B}(\mathcal{H}_{A'})$ iff for any pair of integer sequences $(n_\nu)_{\nu\in\mathbb{N}}$ and $(m_\nu)_{\nu\in\mathbb{N}}$ with $\lim_{\nu\to\infty} n_\nu = \infty$ and $\overline{\lim}_{\nu\to\infty} \frac{m_\nu}{n_\nu} \leq r$ we have*

$$\lim_{\nu\to\infty} \Delta(n_\nu, m_\nu) = 0\,, \tag{5.1}$$

where we have set

$$\Delta(n_\nu, m_\nu) := \inf_{E,D} \|E\, T^{\otimes n_\nu} D - S^{\otimes m_\nu}\|_{cb}\,, \tag{5.2}$$

the infimum taken over all encoding channels E and decoding channels D with suitable domain and range. The channel capacity *$Q(T,S)$ of T with respect to S is defined to be the supremum of all achievable rates. The* quantum capacity *is the special case $Q(T) := Q(T, \mathrm{id}_2)$, with id_2 being the ideal qubit channel.*

There is considerable freedom in this definition, at least for ideal reference channels. We will now briefly discuss a few of the major variations, referring to [KW04] for a complete discussion.

5.2.1 Testing Only One Sequence

At first sight Def. 5.1 appears slightly impractical, since it involves checking an infinite number of pairs of sequences $(n_\nu)_{\nu\in\mathbb{N}}$ and $(m_\nu)_{\nu\in\mathbb{N}}$ when testing a given rate r. But luckily the workload can be substantially reduced: if a coding scheme construction works for a certain pair of integer sequences $(N_\mu)_{\mu\in\mathbb{N}}$, $(M_\mu)_{\mu\in\mathbb{N}}$ such that the rate r is achieved asymptotically, i. e., $\lim_{\mu\to\infty}\frac{M_\mu}{N_\mu} = r$, and the error tends to zero, $\lim_{\mu\to\infty}\Delta(N_\mu, M_\mu) = 0$, then coding works for all such pairs. A proof of this result requires an extension of the given coding scheme to more block sizes. If the given coding scheme is not too sparse, such an extension already follows from basic monotonicity properties of the distance measure Δ.

Obviously, good coding becomes easier the more parallel channels are available for the transmission. Moreover, if a certain coding scheme works for some Hilbert space $\mathcal{H}$, it works at least as well for states supported on a lower-dimensional Hilbert space $\mathcal{H}'$. We thus have

$$\Delta(n+1, m) \leq \Delta(n, m) \leq \Delta(n, m+1) \tag{5.3}$$

for all positive integers $n, m \in \mathbb{N}$. We call a diverging sequence $(N_\mu)_{\mu\in\mathbb{N}}$ *subexponential* if

$$\lim_{\mu\to\infty}\frac{N_{\mu+1}}{N_\mu} = 1\,. \tag{5.4}$$

This covers, for example, all arithmetic sequences, and polynomially growing ones. If the rate r is achieved with vanishing errors along such a subexponential sequence, then the monotonicity properties (5.3) are indeed enough to show that r is an achievable rate in the sense of Def. 5.1. Since this result will prove useful in the investigation of coding schemes for memory channels in Sec. 6.6.2, we reproduce it here from [KW04]:

Proposition 5.2. *Suppose* $\Delta\colon \mathbb{N}\times\mathbb{N} \to \mathbb{R}_+$ *satisfies the monotonicity properties (5.3). Let* $(N_\mu)_{\mu\in\mathbb{N}}$, $(M_\mu)_{\mu\in\mathbb{N}}$ *be a pair of integer sequences such that* $(N_\mu)_{\mu\in\mathbb{N}}$ *is subexponential in the sense of Eq. (5.4), and* $\lim_{\mu\to\infty}\Delta(N_\mu, M_\mu) = 0$. *Then for any pair of integer sequences* $(n_\nu)_{\nu\in\mathbb{N}}$, $(m_\nu)_{\nu\in\mathbb{N}}$ *such that* $\lim_{\nu\to\infty} n_\nu = \infty$ *and*

$$\overline{\lim_{\nu\to\infty}}\,\frac{m_\nu}{n_\nu} < \underline{\lim_{\mu\to\infty}}\,\frac{M_\mu}{N_\mu}\,, \tag{5.5}$$

we have $\lim_{\nu\to\infty}\Delta(n_\nu, m_\nu) = 0$.

Proof: If we have only the monotonicity properties of Δ to draw upon, the way to show that $\Delta(n_\nu, m_\nu) \to 0$ is to find a suitable index $\mu = \mu(\nu)$ for all sufficiently large ν such that $\Delta(n_\nu, m_\nu) \leq \Delta(N_{\mu(\nu)}, M_{\mu(\nu)})$, for which we need

$$n_\nu \geq N_{\mu(\nu)} \qquad \text{and} \qquad m_\nu \leq M_{\mu(\nu)}\,. \tag{5.6}$$

The first inequality we will ensure by defining

$$\mu(\nu) = \min\{\alpha \mid N_\alpha \geq n_\nu\} - 1\,. \tag{5.7}$$

Then

$$N_{\mu(\nu)} \leq n_\nu \leq N_{\mu(\nu)+1} \,, \tag{5.8}$$

and $\lim_\nu \mu(\nu) = \infty$. Hence it remains to show that the second inequality in Eq. (5.6) holds for all sufficiently large ν. We consider

$$\frac{m_\nu}{M_{\mu(\nu)}} = \frac{m_\nu}{n_\nu} \, \frac{n_\nu}{N_{\mu(\nu)+1}} \, \frac{N_{\mu(\nu)+1}}{N_{\mu(\nu)}} \, \frac{N_{\mu(\nu)}}{M_{\mu(\nu)}}. \tag{5.9}$$

In this product the second factor is ≤ 1 by Eq. (5.8), and the third converges to 1 because $(N_\mu)_{\mu\in\mathbb{N}}$ is subexponential. Now pick r_-, r_+ such that strict inequalities

$$\overline{\lim_{\nu\to\infty}} \frac{m_\nu}{n_\nu} < r_- < r_+ < \underline{\lim_{\mu\to\infty}} \frac{M_\mu}{N_\mu} \,, \tag{5.10}$$

hold. Then for all sufficiently large ν the first factor in Eq. (5.9) is $\leq r_-$, and the last factor is $\leq 1/r_+$. Hence the product of the first and last factor in Eq. (5.9) is $\leq r_-/r_+ < 1$. Consequently, Eq. (5.6) holds for all sufficiently large ν, as claimed. ■

This result already covers most sequences $(N_\mu)_{\mu\in\mathbb{N}}, (M_\mu)_{\mu\in\mathbb{N}}$ naturally arising for families of codes. Since Prop. 5.2 draws only on the monotonicity properties (5.3), the result holds true even for the non-ideal reference channels and for all the related capacities investigated in Sec. 5.4.

But what if we only know that the rate r can be attained along some superexponential coding scheme? In this case the basic properties of the error function Δ are not enough for an extension to all sequences [KW04]. However, for a noiseless reference channel $S = \mathrm{id}$ random hash codes can be applied to turn any superexponential coding scheme into a dense protocol [KW02, KW04]. An alternative proof is provided by Devetak's random coding scheme [Dev05], as sketched in Sec. 5.5.4: his protocol works for all block lengths, $N_\mu = \mu$, albeit again for the noiseless reference channel only.

5.2.2 Alternative Error Criteria

We have already explained in Sec. 2.7.1 that the cb-norm difference $\|T - \mathrm{id}\,\|_{cb}$ is by no means the only way to evaluate the distance of a channel T from the noiseless channel id. Another distance measure which has wide currency is the *minimum fidelity*,

$$F_{min}(T_*) := \min_{\|\psi\|=1} f^2(T_*(|\psi\rangle\langle\psi|), \psi) = \min_{\|\psi\|=1} \langle\psi|T(|\psi\rangle\langle\psi|)|\psi\rangle \,, \tag{5.11}$$

where we have used the definition of the fidelity f from Sec. 2.7.2.

Entanglement fidelity is a stabilized version of the minimum fidelity and was introduced by Ben Schumacher in 1996 [Sch96]. It characterizes how well the entanglement between the input states and a reference system not undergoing the noise process is preserved:

for a quantum channel T_*: $\mathcal{B}_*(\mathcal{H}) \to \mathcal{B}_*(\mathcal{H})$ and a quantum state $\varrho \in \mathcal{B}_*(\mathcal{H})$, the entanglement fidelity of ϱ with respect to T_* is given as

$$F_e(\varrho, T_*) := \langle\psi| T_* \otimes \mathrm{id}_{\mathcal{B}_*(\mathcal{H})} (|\psi\rangle\langle\psi|) |\psi\rangle , \tag{5.12}$$

where $|\psi\rangle$ is a purification of ϱ. This quantity can be given the alternative expression [Sch96]

$$F_e(\varrho, T_*) = \sum_i |\mathrm{tr}\varrho\, t_i|^2 , \tag{5.13}$$

where $T_*(\sigma) = \sum_i t_i \sigma\, t_i^*$ is the Kraus decomposition of T (cf. Sec. 2.5). It is clear from Eq. (5.13) that the entanglement fidelity does not depend on the details of the purification process. Obviously, $0 \leq F_e(\varrho, T_*) \leq 1$. Moreover, $F_e(\varrho, T_*) = 1$ implies that T_* is noiseless on the support of ϱ: $T_*|_{\mathrm{supp}(\varrho)} = \mathrm{id}_{\mathrm{supp}(\varrho)}$.

We may then define achievable rates exactly as in Def. 5.1 above, but replace the cb-norm difference in the error function Δ by the usual operator norm $\|\cdot\|_\infty$, or the fidelity measures introduced above. The following proposition, which we again cite from [KW04], shows that all these error criteria can be estimated in terms of each other with dimension-independent bounds, and thus lead to equivalent definitions of quantum channel capacity.

Proposition 5.3. (Equivalent Distance Measures)
Let $\mathcal{H}$ be a Hilbert space, $\dim\mathcal{H} < \infty$, and let T_*: $\mathcal{B}_*(\mathcal{H}) \to \mathcal{B}_*(\mathcal{H})$ be a quantum channel. Then

$$\begin{aligned} 1 - \inf_{\varrho\in\mathcal{B}_*(\mathcal{H})} F_e(\varrho, T_*) &\leq 4\sqrt{1 - F_{min}(T_*)} \\ &\leq 4\sqrt{\|T - \mathrm{id}\,\|_\infty} \\ &\leq 4\sqrt{\|T - \mathrm{id}\,\|_{cb}} \\ &\leq 8\left(1 - \inf_{\varrho\in\mathcal{B}_*(\mathcal{H})} F_e(\varrho, T_*)\right)^{\frac{1}{4}} . \end{aligned} \tag{5.14}$$

The comparison of channels is ultimately based on the comparison of a state to its image, and here the pure states are the worst case. Hence the remarkable insensitivity of the quantum capacity to the choice of the error criterion stems from the observation that the comparison between an arbitrary state and a pure state is rather insensitive to the criterion used. Unfortunately, this equivalence is restricted to capacities with noiseless reference channel $S = \mathrm{id}$. As we have seen in Sec. 3.4, in the vicinity of other (non-ideal) channels, equivalence of the stabilized and unstabilized error criteria may be lost. Dimension-dependent bounds can always be found, but these become useless for capacity purposes, since the dimension of the underlying Hilbert spaces diverges in the large block limit $\nu \to \infty$.

What about the channel fidelity $F_c(T_*)$ and the average fidelity $\overline{F}(T_*)$, as introduced in Sec. 2.7.1? Unlike all the other distances measures discussed above, they do not

involve a maximization over input states, which often makes them much easier to handle and compute. While channel fidelity and average fidelity become equivalent in the large-block limit due to Prop. 2.10, none of them is directly equivalent to the distance measures in Prop. 5.3 above [KW04]. Interestingly, they do still lead to equivalent capacity definitions: given a channel $T_*: \mathcal{B}_*(\mathcal{H}) \to \mathcal{B}_*(\mathcal{H})$ such that $F_c(T_*) \geq 1 - \varepsilon$, it is always possible to find a subspace $\mathcal{H}' \subset \mathcal{H}$ with $\dim \mathcal{H}' \geq \frac{1}{2} \dim \mathcal{H}$ such that $F_{min}(T'_*) \geq 1 - 2\,\varepsilon$ for the compressed channel $T'_* := T_* \,|_{\mathcal{H}'}$ [BKN00, KW04]. In channel capacity applications, the dimension of the Hilbert space $\mathcal{H} \equiv \mathcal{H}(\nu)$ increases exponentially in the large block limit $\nu \to \infty$: $\dim \mathcal{H}(\nu) \sim 2^{n_\nu r}$. Hence, its reduction by a factor $1/2$ does not affect the transmission rate r.

5.2.3 Other Variations

There are a few more variations in the definition of channel capacity that are worth pointing out. In particular, the encoding channels E in Eq. (5.2) may always be restricted to isometric embeddings, $E = V^*(\cdot)V$ with some isometry V.

Instead of requiring the error quantity in Eq. (5.1) to approach zero in the large block limit $\nu \to \infty$, one might feel tempted to impose that the errors vanish completely for some sufficiently large block length, since this is the standard setup in the theory of quantum error correction [KL97, NC00]. While it is true that errors can always be assumed to vanish exponentially in Eq. (5.1), requiring perfect correction may completely change the picture: if a channel has some small positive probability for depolarization, the same also holds for its tensor powers, and no such channel allows the perfect transmission of even one qubit. Hence the capacity for perfect correction will vanish for such channels, while the standard capacity (in accordance with Def. 5.1) will be close to maximal, $Q(T) \approx 1$. The existence of perfect error correcting codes thus only gives lower bounds on the channel capacity, but is not required for a positive transfer rate.

In the other extreme, one might sometimes feel inclined to tolerate (small) finite errors in the transmission. For some $\varepsilon > 0$, we define $Q_\varepsilon(T)$ exactly like the quantum capacity in Def. 5.1, but require only that $\Delta(n_\nu, m_\nu) \leq \varepsilon$ for some sufficiently large ν. Obviously, $Q_\varepsilon(T) \geq Q(T)$ for any quantum channel T. We also have $\lim_{\varepsilon \to 0} Q_\varepsilon(T) = Q(T)$ [KW02, KW04]. In the classical setting even a *strong converse* is known: if $\varepsilon > 0$ is small enough, one cannot achieve bigger rates by allowing small errors, i. e., $C_\varepsilon(T) = C(T)$ [CT91, Win99]. It is still undecided whether an analogous property holds for the quantum capacity $Q(T)$.

5.3 Elementary Properties

In this Section we will briefly discuss some useful elementary properties of the quantum channel capacity $Q(T)$.

The capacity of a composite channel $T_1 \circ T_2$ cannot be bigger than the capacity of the channel with the smallest bandwidth. This in turn suggests that simulating a concatenated channel is in general easier than simulating any of the individual channels. These relations are known as *bottleneck inequalities*:

$$Q(T_1 \circ T_2, S) \leq \min\{Q(T_1, S), Q(T_2, S)\}\,, \tag{5.15}$$

$$Q(T, S_1 \circ S_2) \geq \max\{Q(T, S_1), Q(T, S_2)\}\,. \tag{5.16}$$

Instead of running T_1 and T_2 in succession, we may also run them in parallel. In this case the capacity can be shown to be *superadditive*,

$$Q(T_1 \otimes T_2, S) \geq Q(T_1, S) + Q(T_2, S)\,. \tag{5.17}$$

If both S and one of the channels T_1, T_2 are noiseless, we even have additivity [BDS$^+$96]. However, results on the activation of bound entangled states seem to suggest that the inequality in Eq. (5.17) may be strict for some channels.

Finally, the *two-step coding inequality* tells us that by using an intermediate channel in the coding process we cannot increase the transmission rate:

$$Q(T_1, T_2) \geq Q(T_1, T_3)\; Q(T_3, T_2)\,. \tag{5.18}$$

Applying Eq. (5.18) twice with $T_2 = \mathrm{id}$ and $T_3 = \mathrm{id}$ immediately yields upper and lower bounds on the channel capacity with non-ideal reference channel,

$$\frac{Q(T_1)}{Q(T_2)} \geq Q(T_1, T_2) \geq Q(T_1)\; Q(\mathrm{id}\,, T_2)\,. \tag{5.19}$$

The evaluation of the lower bound in Eq. (5.19) then requires efficient protocols for simulating a noisy channel T_2 with a noiseless resource.

There are few special cases in which the quantum channel capacity can be evaluated relatively easily, the most relevant one being the noiseless channel id_n, where by the subscript n we denote the dimension of the underlying Hilbert space. In this case we have

$$Q(\mathrm{id}_n, \mathrm{id}_m) = \frac{\mathrm{ld}\, n}{\mathrm{ld}\, m}. \tag{5.20}$$

The lower bound $Q(\mathrm{id}_n, \mathrm{id}_m) \geq \mathrm{ld}\, n/\mathrm{ld}\, m$ is immediate from counting dimensions. To establish the upper bound, we use the fact that a noiseless quantum channel cannot simulate itself with a rate exceeding unity: $Q(\mathrm{id}_m, \mathrm{id}_m) \leq 1$. This is just the upper bound we want to prove for the special case $n = m$, and it can be extended to the general case with the help of the two-step coding inequality (5.18): from

$$Q(\mathrm{id}_m, \mathrm{id}_n)\, Q(\mathrm{id}_n, \mathrm{id}_m) \leq Q(\mathrm{id}_m, \mathrm{id}_m) \leq 1 \tag{5.21}$$

we immediately conclude that

$$Q(\mathrm{id}_n, \mathrm{id}_m) \leq \frac{1}{Q(\mathrm{id}_m, \mathrm{id}_n)} \leq \frac{\mathrm{ld}\, n}{\mathrm{ld}\, m}\,, \tag{5.22}$$

where in the last step we have applied the lower bound with the roles of n and m interchanged. This establishes Eq. (5.20). Combining this equation with the two-step coding inequality (5.18), we see that for any channel T

$$Q(T, \mathrm{id}_n) = \frac{\mathrm{ld}\, m}{\mathrm{ld}\, n} Q(T, \mathrm{id}_m), \tag{5.23}$$

which shows that quantum channel capacities relative to noiseless channels of different dimensionality only differ by a constant factor. Fixing the dimensionality of the reference channel then only corresponds to a choice of units. In Def. 5.1 we have followed the usual convention in (quantum) information theory and have chosen the ideal qubit channel id_2 as the standard of reference, hence fixing the unit *bit*.

The upper bound on the capacity of ideal channels can also be obtained from a general upper bound on quantum capacities [HW01], which has the virtue of being easily calculated in many situations. It involves the *transposition map* familiar from Sec. 3.4, which we again denote by Θ, defined as the matrix transposition with respect to some fixed orthonormal basis. We know from Sec. 3.4 that the transposition is positive but not completely positive, and thus does not describe a physical channel. We have $\|\Theta\|_{cb} = d$ for a d-level system [Pau02]. For any channel T and small $\varepsilon > 0$,

$$Q(T) \leq Q_\varepsilon(T) \leq \mathrm{ld}\, \|T \circ \Theta\|_{\mathrm{cb}} =: Q_\Theta(T), \tag{5.24}$$

where Q_ε is the finite error capacity introduced in Sec. 5.2.2 above.

The upper bound $Q_\Theta(T)$ has some remarkable properties, which make it a capacity-like quantity in its own right. For example, it is exactly additive,

$$Q_\Theta(S \otimes T) = Q_\Theta(S) + Q_\Theta(T) \tag{5.25}$$

for any pair S, T of quantum channels. In addition, it satisfies the bottleneck inequality $Q_\Theta(S \circ T) \leq \min\{Q_\Theta(S), Q_\Theta(T)\}$. Moreover, it coincides with the quantum capacity on ideal channels, $Q_\Theta(\mathrm{id}_n) = Q(\mathrm{id}_n) = \mathrm{ld}\, n$, and it vanishes whenever $T \circ \Theta$ is completely positive. In particular, if $\mathrm{id} \otimes T$ maps any entangled state to a state with positive partial transpose, we have $Q_\Theta(T) = 0$.

5.4 Related Capacities

Throughout this Chapter we are primarily concerned with the quantum capacity of a quantum channel. Here we briefly discuss a variety of related capacities, which can be derived from Def. 5.1 by either amending the channel S to be simulated, or allowing Alice and Bob to make use of additional resources. Their interrelations are reviewed in some more detail in [BDS+06].

5.4.1 Classical Capacity

Much interest has been devoted to the hybrid problem of transmitting classical information undistorted over noisy quantum channels. The classical capacity $C(T)$ of a quantum channel $T\colon \mathcal{B}(\mathcal{H}_B) \to \mathcal{B}(\mathcal{H}_A)$ is obtained by choosing the classical ideal one-bit channel rather than the one-qubit channel as the standard of reference in Def. 5.1. Encoding channels E and decoding channels D are then restricted to preparations and measurements, respectively. From our discussion in Sec. 2.4 we know that every such encoding channel $E\colon \mathcal{B}(\mathcal{H}_A) \to \mathcal{C}_X$ is of the form

$$E(A) = \sum_{x=1}^{|X|} \varrho_x(A)\, |x\rangle\langle x|\,, \tag{5.26}$$

where $\{\varrho_x\}_{x=1}^{|X|} \subset \mathcal{B}_*(\mathcal{H}_A)$ is a collection of quantum states. The decoding operations $D\colon \mathcal{C}_X \to \mathcal{B}(\mathcal{H}_B)$ are given in terms of a corresponding positive operator-valued measure $\{M_y\}_{y=1}^{|X|} \subset \mathcal{B}(\mathcal{H}_B)$ in the sense of Def. 2.1,

$$D(f) = \sum_{y=1}^{|X|} f_y\, M_y \tag{5.27}$$

for all functions $f \equiv \sum_y f_y\, |y\rangle\langle y|$ on the finite set X. Defining the classical transition matrix $R_{xy} := \varrho_x\big(T(M_y)\big) = \operatorname{tr} T_*(\varrho_x)\, M_y$, the concatenated channel $E \circ T \circ D$ now takes on the form

$$R(f) := E \circ T \circ D\,(f) = \sum_{x,y=1}^{|X|} f_y\, R_{xy}\, |x\rangle\langle x|\,. \tag{5.28}$$

Def. 5.1 requires that we compare the classical channel $R\colon \mathcal{C}_X \to \mathcal{C}_X$ with the noiseless classical channel id_X. We know from Sec. 2.7.1 that cb-norm and operator norm coincide for Abelian algebras. Hence, the cb-norm distance $\|R - \mathrm{id}_X\|_{cb}$ can be evaluated as follows,

$$\begin{aligned}
\|R - \mathrm{id}\,\|_{cb} &= \|R - \mathrm{id}_X\|_\infty \\
&= \sup_{\|f\|_\infty \le 1} \|R(f) - f\|_\infty \\
&= \sup_{x=1,\ldots,|X|} \sup_{\|f\|_\infty \le 1} |(Rf)(x) - f_x| \\
&= \sup_{x=1,\ldots,|X|} \sup_{\|f\|_\infty \le 1} |\sum_y (R_{xy} - \delta_{xy})\, f_y| \\
&= \sup_{x=1,\ldots,|X|} \sum_y |R_{xy} - \delta_{xy}| \\
&= \sup_{x=1,\ldots,|X|} \{|R_{xx} - 1| \,+ \sum_{y \ne x} R_{xy}\} \\
&= 2 \sup_{x=1,\ldots,|X|} \{1 - R_{xx}\}\,,
\end{aligned} \tag{5.29}$$

where in the third to last step the supremum with respect to f is evaluated by choosing $f_y := \mathrm{sign}\,(R_{xy} - \delta_{xy})$ for all $y = 1, ..., |X|$. In the last step we have made use of the normalization of transition probabilities, $\sum_{y=1}^{|X|} R_{xy} = 1$ for all $x = 1, ..., |X|$. For classical channels, the cb-norm distance thus has an immediate interpretation as (twice) the maximal error probability. Dropping the irrelevant factor, Def. 5.1 then leads to the following definition for the classical channel capacity of a quantum channel:

Definition 5.4. (Classical Channel Capacity)
Let $T\colon \mathcal{B}(\mathcal{H}_B) \to \mathcal{B}(\mathcal{H}_A)$ *be a quantum channel. A positive number* r *is called an* achievable rate for classical information transfer *if for every* $\varepsilon > 0$ *there exists a positive integer* $N_\varepsilon \in \mathbb{N}$ *such that for every* $n \geq N_\varepsilon$ *we may find a code book with* $\nu := \lfloor 2^{nr} \rfloor$ *codewords* $\{\varrho_j\}_{j=1}^{\nu} \subset \mathcal{B}_*(\mathcal{H}_A)^{\otimes n}$ *and a corresponding POVM* $\{M_j\}_{j=1}^{\nu} \subset \mathcal{B}(\mathcal{H}_B)^{\otimes n}$ *such that*

$$\mathrm{tr}\, T_*^{\otimes n}(\varrho_j)\, M_j \;\geq\; 1 - \varepsilon \qquad \forall\; j = 1, ..., \nu\,. \tag{5.30}$$

The classical channel capacity $C(T)$ *of* T *is again defined as the supremum of all achievable rates.*

In Sec. 5.5.4 we will introduce the *private classical capacity* $C^p(T)$ — a closely related variant in which the coding scheme has to satisfy the additional constraint that no information is released to a potential eavesdropper, who is assumed to hold a purification of the channel T. Obviously, $C(T) \geq C^p(T)$.

Since a quantum channel can always be employed to send classical information, we also have the inequality $C(T) \geq Q(T)$. There are obviously examples in which this inequality is strict: the *entanglement-breaking channel* $T_*(\varrho) = \sum_j \langle j|\varrho|j\rangle\, |j\rangle\langle j|$ is composed of a measurement in the orthonormal basis $\{|j\rangle\}_j$, followed by a preparation of the corresponding basis states $|j\rangle\langle j|$. It destroys all the entanglement between the sender and a reference system, implying $Q(T) = 0$. Yet all the basis states $|j\rangle\langle j|$ are transmitted undistorted, which is enough to guarantee that $C(T) = 1$.

5.4.2 Enhanced Capacities

Superdense coding [BW92] and teleportation [BBC$^+$93] impressively demonstrate that entanglement is a powerful resource for information transfer. It doubles the classical channel capacity of a noiseless channel, and it allows to send quantum information over purely classical channels. We define the *entanglement-assisted quantum capacity* $Q^E(T)$ exactly like $Q(T)$ in Def. 5.1, but allow Alice and Bob to draw on an unlimited supply of shared ebits as an additional resource for the channel coding. The entanglement-assisted classical capacity $C^E(T)$ is defined completely analogously, but again we only require the simulation of a classical bit channel. Superdense coding and teleportation imply that $C^E(T) = 2\, Q^E(T)$ for any quantum channel T.

Surprisingly, the *entanglement-assisted capacities* are often much better behaved than their unassisted counterparts. As we will see in Sec. 5.5.2, unlike the classical and quantum capacities proper they are relatively easy to calculate using finite optimization procedures [BSS$^+$99, BSS$^+$02]. Moreover, there has recently been significant progress in understanding the simulation rates for non-ideal channels in this scenario [Hay06^c].

The quantum channel capacity is unaffected by entanglement-breaking side channels [KW04]. In particular, classical forward communication alone cannot enhance it. However, unlike in the purely classical case both the quantum and classical channel capacity (but not the entanglement-assisted capacity) may increase under classical feedback [Bow04^b, Bow05].

5.5 Coding Theorems

Computing channel capacities straight from Def. 5.1 is a tricky business. It involves optimization in systems of asymptotically many tensor factors, and can only be performed in special cases, like the noiseless channels discussed in Sec. 5.3. *Coding Theorems* aspire to reduce this problem to an optimization over a low-dimensional space. They usually come in two parts: the *converse* provides an upper bound on the channel capacity (typically in terms of some entropic expression), while the *direct* part consists of a coding scheme that attains this bound. The prototype for all such results is Shannon's celebrated noisy channel coding theorem [Sha48, Ash90, CT91], which proves that the classical capacity of a classical noisy channel can be obtained from a maximization of the *mutual information* over all joint input-output distributions.

In this Section we will review the known extensions of Shannon's channel coding theorem to the quantum world. Again, the focus will be mostly on the quantum channel capacity proper, but we will briefly comment on the classical capacity and the entanglement-assisted capacities first.

5.5.1 Classical Channel Capacity: the HSW Theorem

The classical capacity of a quantum channel $T\colon \mathcal{B}(\mathcal{H}_B) \to \mathcal{B}(\mathcal{H}_A)$, as defined in Sec. 5.4.1 above, can be expressed in terms of the *Holevo bound*,

$$\chi(T_*, \{p_i, \varrho_i\}) := H\big(\sum_i p_i\, T_*(\varrho_i)\big) - \sum_i p_i\, H\big(T_*(\varrho_i)\big)\,, \tag{5.31}$$

(cf. App. A). Holevo [Hol73] was able to show back in 1973 that the regularized Holevo bound provides an upper bound on the classical channel capacity when the code book consists of the quantum ensemble $\{p_i, \varrho_i\}$. Attainability of this bound was proved much later by Holevo [Hol98], and independently by Schumacher and Westmoreland [SW97].

In honor of these researchers, the coding theorem for the classical channel capacity is usually called the HSW theorem.

Theorem 5.5. (HSW)
For every quantum channel $T\colon \mathcal{B}(\mathcal{H}_B) \to \mathcal{B}(\mathcal{H}_A)$,

$$C(T) = \lim_{n\to\infty} \frac{1}{n} \max_{\{p_i,\varrho_i\}} \chi(T_*^{\otimes n}, \{p_i, \varrho_i\}). \tag{5.32}$$

In contrast to Shannon's noisy coding theorem, the evaluation of the capacity formula Eq. (5.32) still requires an optimization in spaces of asymptotically increasing dimensions. It is a long-standing open conjecture that entangled input states cannot enhance the transmission rate for classical information, which would allow to remove the regularization from Eq. (5.32). The additivity conjecture can be traced back to [BFS97] and has wide implications for other problems in the field [Sho04, Pom03]. It is presently considered one of the most eminent open questions in quantum information theory. We refer to [Hol06^b] for a well-informed account of the present status.

5.5.2 Entanglement-Assisted Capacities

The entanglement-assisted capacities introduced in Sec. 5.4.2 have a nice representation in terms of a single-letter entropic formula. Bennett and coworkers [BSS$^+$99, BSS$^+$02] have shown that the classical entanglement-assisted capacity $C^E(T)$ of a quantum channel $T\colon \mathcal{B}(\mathcal{H}_B) \to \mathcal{B}(\mathcal{H}_A)$ simply equals the optimized quantum mutual information,

$$H(T_*, \varrho) := H\big(\varrho\big) + H\big(T_*(\varrho)\big) - H\big(T_* \otimes \mathrm{id}\,(|\psi_\varrho\rangle\langle\psi_\varrho|)\big)\,, \tag{5.33}$$

where $\psi_\varrho \in \mathcal{H}_A \otimes \mathcal{H}_{A'}$ is a purification of the density operator $\varrho \in \mathcal{B}_*(\mathcal{H}_A)$ (cf. App. A). As explained in Sec. 5.4.2, teleportation and superdense coding immediately imply that this is just twice the quantum entanglement-assisted capacity $Q^E(T)$, resulting in the following

Theorem 5.6. (Entanglement-Assisted Capacities)
For every quantum channel $T\colon \mathcal{B}(\mathcal{H}_B) \to \mathcal{B}(\mathcal{H}_A)$,

$$C^E(T) = \max_{\varrho} H(T_*, \varrho) = 2\, Q^E(T)\,. \tag{5.34}$$

We refer to Hayden's review article [Hay06^c] for a more detailed account and recent progress in the understanding of entanglement-assisted capacities.

5.5.3 Quantum Channel Capacity

For the quantum channel capacity the relevant entropic quantity is the so-called *coherent information*,

$$I_c(T_*, \varrho) := H\big(T_*(\varrho)\big) - H\big(T_* \otimes \mathrm{id}\,(|\psi_\varrho\rangle\langle\psi_\varrho|)\big)\,, \tag{5.35}$$

where $|\psi\rangle_\varrho \in \mathcal{H}_A \otimes \mathcal{H}_{A'}$ is again a purification of the density operator $\varrho \in \mathcal{B}_*(\mathcal{H}_A)$. We know from App. A that $I_c(T_*, \varrho)$ does not increase under quantum operations, which immediately implies that the regularized coherent information upper bounds the quantum channel capacity: if Alice and Bob have a coding scheme for the channel T with capacity $Q(T)$, n channel uses allow them to share a maximally entangled state of size $\sim \exp_2 n\,Q(T)$. The coherent information of this state equals $\sim n\,Q(T)$, and was no smaller prior to Bob's decoding operation D. Recently Devetak developed a coding scheme [Dev05] to show that this bound is in fact attainable. Different proofs were outlined by Lloyd [Llo97] and Shor [Sho02].

Theorem 5.7. (Quantum Channel Capacity)
For every quantum channel $T\colon \mathcal{B}(\mathcal{H}_B) \to \mathcal{B}(\mathcal{H}_A)$,

$$Q(T) = \lim_{n\to\infty} \frac{1}{n} \max_{\varrho} I_c\big(T_*^{\otimes n}, \varrho\big)\,. \tag{5.36}$$

Unlike the classical or quantum mutual information, coherent information is strictly superadditive for some channels [DSS98]. Hence taking the limit $n \to \infty$ in Eq. (5.36) is indeed required, and in general the evaluation of the capacity formula (5.36) still demands the solution of asymptotically large variational problems. This should be contrasted with the coding theorems for the entanglement-assisted capacities in Sec. 5.5.2 (where a simple non-regularized coding theorem is known to hold) and for the classical information capacity in Sec. 5.5.1 (where additivity is conjectured but not proved). Even the maximization of the single-shot coherent information $I_c(T_*, \varrho)$ appears to be a difficult optimization problem, since this quantity is neither convex nor concave and may have multiple local maxima [Sho03]. Thus even for simple-looking systems like the qubit depolarizing channel, so far we only have upper and lower bounds on the quantum channel capacity, but do not yet know how to compute its exact value.

5.5.4 Private Classical and Quantum Coding

We now sketch Devetak's proof of Th. 5.7, assuming only some familiarity with Holevo-Schumacher-Westmoreland (HSW) random codes for the classical channel capacity [Hol98, SW97, NC00]. Along the way we will also obtain a coding theorem for the *private* classical capacity of a quantum channel.

As shown in Sec 3.3, Stinespring's dilation theorem implies that a noiseless quantum channel provides perfect security against eavesdropping. This is one of the characteristic traits of quantum mechanics and lies at the heart of quantum cryptography. In his proof, Devetak showed a way to turn this around and upgrade coding schemes for private classical information to quantum channel codes.

The relation between quantum information transfer over a channel $T\colon \mathcal{B}(\mathcal{H}_B) \to \mathcal{B}(\mathcal{H}_A)$ and privacy against eavesdropping is best understood in terms of the *complementary*

channel $T_E\colon \mathcal{B}(\mathcal{H}_E) \to \mathcal{B}(\mathcal{H}_A)$. As explained in Sec. 3.3, T_E arises from a given Stinespring isometry $V\colon \mathcal{H}_A \to \mathcal{H}_B \otimes \mathcal{H}_E$ of $T \equiv T_B$ by interchanging the roles of the output system $\mathcal{B}(\mathcal{H}_B)$ and the environment $\mathcal{B}(\mathcal{H}_E)$:

$$T_B(B) = V^*(B \otimes \mathbb{1}_E)V \qquad \Longleftrightarrow \qquad T_E(E) = V\,(\mathbb{1}_B \otimes E)V^* \tag{5.37}$$

for all observables $B \in \mathcal{B}(\mathcal{H}_B)$ and $E \in \mathcal{B}(\mathcal{H}_E)$. The channel T_E describes the information flow into the environment $\mathcal{B}(\mathcal{H}_E)$, a system we assume to be under complete control of a potential eavesdropper, Eve say. The setup for private classical information transfer (including the definition of rates and capacity $C^p(T)$) is then exactly the same as for the classical channel capacity $C(T)$ in Sec. 5.4.1, but the protocols now have to satisfy the additional requirement that T_E releases (almost) no information to the environment. This can be achieved by randomizing over $\nu_E \sim \exp_2 n\,\chi(T_{E*}, \{p_i, \varrho_i\})$ code words of a standard HSW code of total size $\sim \exp_2 n\,\chi(T_{B*}, \{p_i, \varrho_i\})$, where $\{p_i, \varrho_i\}$ is the quantum ensemble from which a set of random code words $\{\sigma_{k,l}\}_{k=1,l=1}^{\nu_B,\nu_E}$ is generated. The appearance of the *Holevo bound* Eq. (5.31) in the dimension of both these code spaces can be understood from the size of the relevant typical subspaces [DW04^b].

The randomization guarantees that the remaining $\nu_B \sim \exp_2 n\,\big(\chi(T_{B*}) - \chi(T_{E*})\big)$ code words are almost indistinguishable to Eve:

$$\Big\| \frac{1}{\nu_E} \sum_{l=1}^{\nu_E} T_{E*}^{\otimes n}\,(\sigma_{kl} - \sigma_{jl}) \Big\|_1 \leq \varepsilon \qquad \forall\ j,k = 1, ..., \nu_B. \tag{5.38}$$

The net transfer rate for private classical information is then $r \sim \big(\chi(T_{B*}) - \chi(T_{E*})\big)$, which is just the total transfer rate for the channel Alice $\to$ Bob reduced by the transfer rate Alice $\to$ Eve. In summary, Devetak's proof leads to the following coding theorem for the private classical capacity:

Theorem 5.8. (Private Channel Capacity)
For every quantum channel $T\colon \mathcal{B}(\mathcal{H}_B) \to \mathcal{B}(\mathcal{H}_A)$,

$$C^P(T) = \lim_{n\to\infty} \frac{1}{n} \max_{\{p_i,\varrho_i\}} \Big\{ \chi(T^{\otimes n}, \{p_i, \varrho_i\}) - \chi(T_E^{\otimes n}, \{p_i, \varrho_i\}) \Big\}. \tag{5.39}$$

So far we have sketched a protocol for private classical information transfer. Remarkably, if $\varrho = \sum_i p_i\,|\psi_i\rangle\langle\psi_i|$ is a decomposition of $\varrho \in \mathcal{B}_*(\mathcal{H}_A)$ into *pure* states, the private transfer rate exactly equals the coherent information,

$$I_c(T_{B*}, \varrho) = H\big(T_{B*}(\varrho)\big) - H\big(T_{E*}(\varrho)\big) = \chi(T_{B*}) - \chi(T_{E*})\,. \tag{5.40}$$

The so-called *entropy exchange* $H\big(T_{E*}(\varrho)\big) = H\big(T_{B*} \otimes \mathrm{id}\,(|\psi_\varrho\rangle\langle\psi_\varrho|)\big)$ quantifies to what extent a formerly pure ancilla state becomes mixed via interaction with the signal states. Eq. (5.40) then nicely reflects the intuition that for high rate quantum information transfer the signal states should not entangle too much with the environment. In fact,

for an almost noiseless channel the entropy exchange nearly vanishes, and the optimized coherent information almost attains the maximal value 1, while for nearly depolarizing channels we have $I_c(T_{B*}, \varrho) \approx -H(\varrho) \leq 0$.

Devetak's *coherentification* allows to pass from the transmission of classical messages to the transmission of coherent superpositions. The same technique has also been applied to upgrade secret key distillation protocols to entanglement distillation protocols, and provides a unified perspective of the secret classical resources and their quantum counterparts [DW04^a, DW04^b, DHW04].

In order to transfer quantum information faithfully, Alice will only need to send one half of a maximally entangled state of dimensionality $\nu_B \sim \exp_2 n\, I_c(T_{B*}, \varrho)$. The standard teleportation protocol [BBC$^+$93] then allows her to transfer arbitrary quantum states from a subspace of that size. Teleportation requires classical forward communication, but we have seen in Sec. 5.4.2 that this cannot increase the quantum channel capacity.

Given a set of pure state code words $\{|\varphi_{kl}\rangle\}_{k=1,l=1}^{\nu_B,\nu_E}$ of a private classical information protocol, for entanglement transfer Alice prepares the input state

$$|\Phi\rangle_{A'A} = \frac{1}{\sqrt{\nu_B}} \sum_{k=1}^{\nu_B} |k\rangle_{A'} \otimes \frac{1}{\sqrt{\nu_E}} \sum_{l=1}^{\nu_E} |\varphi_{kl}\rangle_A \,, \tag{5.41}$$

where $\mathcal{H}_{A'} \simeq \mathcal{H}_A$ denotes a reference system that Alice keeps in her lab. On his share of the resulting output state $|\Phi'\rangle_{A'BE}$ Bob will then employ the corresponding measurement operators $\{M_{kl}\}_{k,l=1}^{\nu_B,\nu_E}$ to implement the *coherent* measurement

$$V_M\, |\varphi\rangle_B := \sum_{k,l} \sqrt{M_{kl}}\, |\varphi\rangle_B \otimes |kl\rangle_{B_1B_2} \,, \tag{5.42}$$

which places the measurement outcomes into some reference system $\mathcal{H}_{B_1} \otimes \mathcal{H}_{B_2}$. Any measurement which identifies the output with high probability only slightly disturbs the output state, and thus Bob's coherent measurement leaves the total system in some approximation of the state

$$|\Phi''\rangle = \frac{1}{\sqrt{\nu_B \nu_E}} \sum_{k,l}^{\nu_B,\nu_E} |k\rangle_{A'}\, |k\rangle_{B_1}\, |l\rangle_{B_2}\, |\varphi'_{kl}\rangle_{BE} \,, \tag{5.43}$$

in which Eve and Bob are still entangled. A completely depolarizing channel T_E would directly yield a factorized output state in $\mathcal{B}_*(\mathcal{H}_B) \otimes \mathcal{B}_*(\mathcal{H}_E)$ here. Although the randomization in Eq. (5.38) does not necessarily result in complete depolarization, there is a controlled unitary operation which Bob may apply to effectively decouple Eve's system, resulting in the output state $\sim \frac{1}{\sqrt{\nu_B}} \sum_k |kk\rangle_{A'B_1} \otimes \sigma_E$ for some state $\sigma_E \in \mathcal{B}_*(\mathcal{H}_E)$. For Alice and Bob this is now the maximally entangled state of size $\nu_B \sim \exp_2 n I_c(T_{B*}, \varrho)$ required for teleportation. The direct part of the capacity theorem then follows by applying the above coding scheme to large blocks and maximizing over (pure) input ensembles, concluding the proof.

Devetak's proof of the coding theorem seems to indicate that the private classical capacity $C^p(T)$ equals the quantum capacity $Q(T)$ for every quantum channel T. However, for the coherentification protocol we have restricted the private coding schemes to pure state input ensembles, and thus we can only conclude that $Q(T) \leq C_p(T)$. The existence of bound-entangled states with positive one-way distillable secret key rate [HPH$^+$05] implies that this inequality can be strict. We have seen in Sec. 3.3 that a general procedure does exist to retrieve (almost) all the information from the output of a noisy quantum channel that releases (almost) no information to the environment. But this requires a stronger form of privacy than Eq. (5.38).

Chapter 6

Quantum Channels with Memory

In this Chapter we will present a general model for quantum channels with memory, and we will show that this model is sufficiently general to encompass all *causal* automata: any quantum process in which the outputs up to some time t do not depend on inputs at times $t' > t$ can be decomposed into concatenated memory channels. We then examine and present different physical setups in which channels with memory may be operated for the transfer of (private) classical and quantum information. These include setups in which either the receiver or a malicious third party have control of the initializing memory. We introduce classical and quantum channel capacities for these settings, and give several examples to show that they may or may not coincide. Entropic upper bounds on the various channel capacities are given. For *forgetful* quantum channels, in which the effect of the initializing memory dies out as time increases, coding theorems are presented to show that these bounds may be saturated. Forgetful quantum channels are shown to be open and dense in the set of quantum memory channels.

Most of the results reported here have appeared in a joint paper [KW05] with R. F. Werner.

6.1 Introduction

Until now most of the work on quantum channels has concentrated on *memoryless* channels, which are characterized by the requirement that successive channel inputs for some quantum channel S are acted on independently. As explained in Ch. 5, mathematically this means that messages of n symbols are processed by the tensor product channel $S^{\otimes n}$.

However, in many real-world applications the assumption of having uncorrelated noise channels cannot be justified, and *memory effects* need to be taken into account. It thus seems desirable to extend the theory of quantum channels to encompass memory

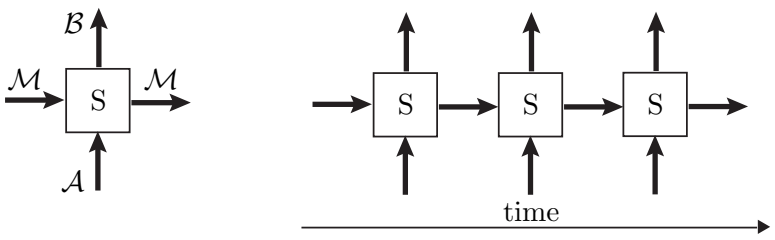

Figure 6.1: Left: *A quantum memory channel with input register $\mathcal{A}$, output register $\mathcal{B}$, and memory system $\mathcal{M}$.* — Right: *A threefold concatenation S_3 of memory channels, with time running from left to right, and coded information running from bottom to top.*

effects, and to create a common framework in which experiments with both correlated and uncorrelated noise can be naturally described. In fact, such a framework is already necessary for estimates on almost memoryless channels, for instance when assessing whether a particular system can arguably be modelled as a memoryless channel. Such a unified framework will be presented in this Chapter, and it will be shown how this model can be applied to the description of different information processing tasks, such as (private) classical and quantum information transfer.

6.1.1 Outline and Overview

We start our description with a general model for quantum channels with memory. In addition to Alice's input register $\mathcal{A}$ and Bob's output register $\mathcal{B}$, such a channel has an additional memory input and an additional memory output, denoted by $\mathcal{M}$ (cf. Fig. 6.1, left). Long messages with n signal states will then be processed by subsequent application of these memory channels, resulting in the concatenated channel S_n depicted in Fig. 6.1 (right). This picture will be turned into a rigorous definition in Sec. 6.2.1.

In such a setup, the memory system is passed on from one application of the channel to the next, and introduces (quantum or classical) correlations between consecutive signal states. If no memory system is present, the concatenated channel will simply be a product channel, bringing us back to the memoryless realm in which consecutive signal states are acted on independently.

This model marks a *constructive approach* to quantum channels with memory. It is certainly the appropriate framework when the physical realization of the memory $\mathcal{M}$ is known. However, in many applications of information theory only the input-output behavior of a channel is of interest. From this point of view the memory would be part of the internal workings of the channel, and would not be made part of the description. We call this way of describing channels the *axiomatic approach.* It takes a channel as a transformation turning infinite strings of input systems to infinite strings of outputs,

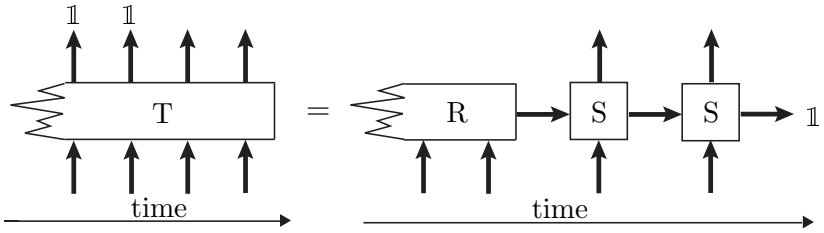

Figure 6.2: *By the* structure theorem*, a causal automaton T can be decomposed into a chain of concatenated memory channels S plus some input initializer R. Evaluation with the identity operator $\mathbb{1}$ means that the corresponding output is ignored.*

with only two basic assumptions: translational invariance and the condition of *causality*. Outputs up to some time t do not depend on inputs at times $t' > t$. In the classical theory, such channels are sometimes called *non-anticipatory*. It is clear from Fig. 6.1 that a channel with memory automatically satisfies this causality condition.

Taking a causal channel and representing it as a channel with memory amounts to reconstructing a model of the channel and its internal memory states and dynamics. This is a highly non-trivial task, even in the classical case. However, a formal reconstruction can always be given. This is what we call the *structure theorem* for causal channels, as illustrated in Fig. 6.2. A rigorous version will be given as Th. 6.10 in Sec. 6.3. In general, it produces not only the channel step operator S, but also a map R defining the influence of input states in the remote past on the memory. Intuitively, however, such a map is often not needed, because memory effects decrease in time. A similar condition is needed for passing from the constructive approach of channels with memory to causal input-output channels: Since the constructive approach allows one to choose the initial memory state, output states in general depend on this choice, and in general this influence will depend on the time after initialization. So in order to get a time translation invariant channel without such dependence, the channel S must lose the initialization information. We call S *forgetful* if outputs at a large time t depend only weakly on the memory initialization at time zero, in a sense made precise in Sec. 6.4. For forgetful channels, memory effects will be shown to decrease even exponentially.

Not every channel is forgetful. The prime counterexample is a channel with a *global classical switch* discussed in Sec. 6.2.3. The memory in this case is a classical bit, left unchanged by S, but determining which of two memoryless channels S_0, S_1 is applied to the input at each time step. However, we will show in Sec. 6.4 that generic memory channels are in fact forgetful, in the sense that every non-forgetful quantum channel can be approximated by a forgetful channel to arbitrary degree of accuracy. In addition, for every forgetful quantum channel we may find a finite-size neighborhood in which all channels are likewise forgetful. In mathematical terms, forgetful quantum channels

are both *open* and *dense* in the set of quantum memory channels.

For quantum channels with memory, capacity can be defined along the lines familiar from the memoryless setting reviewed in Ch. 5, both for the transmission of classical and quantum information. For product as well as memory channels, channel capacity expresses quantitatively how well a given channel S can simulate a noiseless qubit (or bit) channel: roughly speaking, it is the maximal number of ideal qubit (resp. bit) transmissions per use of the channel, taken in the limit of long messages and using encoding and decoding schemes asymptotically eliminating all errors. The concept is illustrated in Fig. 6.3. This figure should be compared to Fig. 5.1, which illustrates the capacity of a memoryless channel.

When trying to send information through a concatenated memory channel, unlike in the memoryless case we also have to specify how to handle the initial and final memory state. In particular, we may distinguish between setups in which Alice can access the initial memory input state and may use it for the encoding procedure, and setups in which a malicious third party (Eve, say) controls the initial memory input, and by her choice of the input state will try to prevent Alice and Bob from communicating over the channel. Likewise, we may consider setups in which either Bob or Eve control the final memory output. These distinctions will be made precise in Sec. 6.2.2. They lead to slight variations in the notion of capacity, and in Sec. 6.2.3 we will present several examples to show that the resulting capacities may or may not coincide. In particular, for channels with only one Kraus operator, all these capacities are the same, and equal the capacity of the ideal channel (cf. Sec. 6.2.4).

The various capacities can be bounded from above both in terms of the capacity of memoryless channels and in terms of entropic expressions. Some of these bounds will be presented. In particular, the standard mutual information and coherent information bounds familiar from the memoryless setting easily extend to memory channels (cf. Sec. 6.6.1).

Forgetful channels are, in a sense to be specified in Sec. 6.4, close to memoryless channels. As such, they play a central role not only as the bridge between the axiomatic and the constructive approach to quantum memory channels and as generic examples, but also connect them to the memoryless realm. In Sec. 6.6.2, we will explain how the standard random coding techniques familiar from the memoryless setting can be modified to saturate the entropic upper bounds on the channel capacity for forgetful channels, leading to coding theorems for (private) classical and quantum information transfer for this very important class of memory channels.

We conclude with a Summary and Outlook. Appendix B contains some mathematical background relevant to the description of infinite-dimensional quantum systems, insofar as it is essential to the understanding of the Structure Theorem.

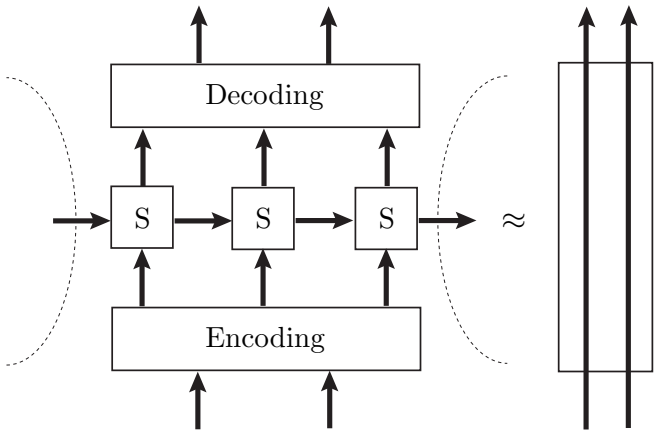

Figure 6.3: *Two signal states are encoded into three input registers, sent through the concatenated memory channel, and then decoded into two output states. If the overall channel is (in some sense to be specified in Sec. 6.2.2) close to the ideal channel on two inputs, the transmission rate of the above scheme is* $\frac{2}{3}$*. Capacity is the largest such rate, in the limit of long messages and optimal encoding and decoding. In the above setup, the initial memory input can be thought of as being controlled by either the sender or a malicious third party. Similarly, the receiver may or may not be able to read out the final memory state.*

6.1.2 Model Systems and Related Work

Quantum channels which naturally acquire a memory are abundant in all branches of quantum information processing:

Recently, an unmodulated spin chain has been proposed as a model for short distance quantum communication [Bos03, BB05, CDE$^+$04, CDD$^+$05]. In such a scheme, the state to be communicated over the channel is placed on one of the spins of the chain, propagates for a specific amount of time, and is then received at a distant spin of the chain (cf. Fig. 6.4). When viewed as a model for quantum communication, it is generally assumed that a reset of the spin chain occurs after each signal [GF05], for example by applying an external magnetic field, resulting in a memoryless channel. However, a continuous operation without reset may lead to higher transmission rates, and corresponds to a quantum channel with memory. Another model of a quantum channel with memory is the so-called *one-atom maser* or *micromaser* [MWM85, VBW$^+$00]. In such a device, excited atoms interact with the photon field inside a high-quality optical cavity, as depicted in Fig. 6.5. If the photons inside the cavity have sufficiently long lifetime, atoms entering the cavity will feel the effect of the predecessors, introducing correlations between consecutive signal states.

Apparently, the first model of a quantum channel with memory was introduced by

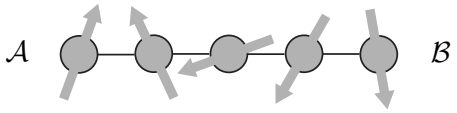

Figure 6.4: *An unmodulated spin chain as a quantum channel with memory: Alice places the input signal on the first spin of the chain and lets it propagate to Bob, who controls the spin at the opposite end of the chain.*

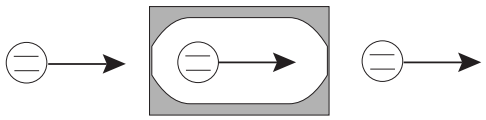

Figure 6.5: *In a* micromaser*, a stream of two-level atoms is injected into a high-quality superconducting cavity. The field modes introduce correlations between consecutive atoms.*

Macchiavello *et al.* in 2001: they gave an example of a qubit channel with Markovian correlated noise [MP02, MPV04] in which entangled input states may increase the transmission rate for classical information. These results have recently been extended to some bosonic Gaussian channels [CCM$^+$05, RSG$^+$05]. Such an effect has been demonstrated experimentally for optical fiber channels with fluctuating birefringence, in which consecutive light pulses undergo strongly correlated polarization transformation [BDB04, BDW$^+$04]. (Whether such examples exist in the *memoryless* setting is still an open question, and presently considered one of the most eminent open problems of quantum information theory, with wide implications for other problems in the field [Sho04, Pom03].)

Subsequently, the study of quantum channels with memory has largely been confined to channels with Markovian correlated noise (cf. [Ham02, BM04] and references therein). A Lindbladian approach to memory channels has been taken by Daffer et al. [DWM03, DWC$^+$04]. Upper bounds on the classical capacity for a more general class of channels have been given recently by Bowen *et al.* [BDM05].

All the memory channels discussed in this Section are causal quantum channels, and hence the structure theorem applies. A completely different approach has been taken by Hayashi and Nagaoka [HN03], who refrain from imposing any structural assumption on the quantum channels they consider, and apply the information-spectrum method to obtain a coding theorem for the classical product state capacity, following work by Verdú and Han [VH94] on classical channels with memory. A more detailed account of the quantum information spectrum techniques will be given in Ch. 7.

We refer to Verdú's overview paper [Ver98] and the Gray-Davisson collection [GD77] for more information on memory channels in the purely classical setting.

6.2 Channels with Memory

In Sec. 6.2.1 we turn the concept of memory channels into a rigorous definition. We then focus on capacities for classical and quantum information transfer in Sec. 6.2.2.

Some instructive examples will serve to illustrate the various capacities in Sec. 6.2.3, before we turn to pure channels in Sec. 6.2.4.

6.2.1 The Constructive Approach

A relatively simple (yet surprisingly general, cf. Sec. 6.3 below) model to describe channels with correlated noise consists of a quantum channel which, in addition to Alice's input register system $\mathcal{H}_A$ and Bob's output register system $\mathcal{H}_B$ has an additional memory input $\mathcal{H}_M$ and an additional memory output $\mathcal{H}_{M'}$. (Since the smaller of the two Hilbert spaces $\mathcal{H}_M$, $\mathcal{H}_{M'}$ can always be thought of as being embedded in the larger one, in the following we will assume without loss that $\mathcal{H}_M = \mathcal{H}_{M'}$.) A *quantum channel with memory* (or, for short, *memory channel*) is then represented (in Heisenberg picture) as a completely positive and unital map $S\colon \mathcal{B}(\mathcal{H}_B) \otimes \mathcal{B}(\mathcal{H}_M) \to \mathcal{B}(\mathcal{H}_M) \otimes \mathcal{B}(\mathcal{H}_A)$. Often we will abbreviate $\mathcal{B}(\mathcal{H}_A)$ to $\mathcal{A}$, and similarly for $\mathcal{B}(\mathcal{H}_B)$ and $\mathcal{B}(\mathcal{H}_M)$. Long messages with $n \in \mathbb{N}$ signal states will now be processed by subsequent application of memory channels, resulting in the concatenated channel $S_n\colon \mathcal{B}^{\otimes n} \otimes \mathcal{M} \to \mathcal{M} \otimes \mathcal{A}^{\otimes n}$ given as follows (see Fig. 6.1):

$$S_n = \left(S \otimes \mathrm{id}_{\mathcal{A}}^{\otimes n-1}\right) \circ \ldots \circ \left(\mathrm{id}_{\mathcal{B}}^{\otimes n-2} \otimes S \otimes \mathrm{id}_{\mathcal{A}}\right) \circ \left(\mathrm{id}_{\mathcal{B}}^{n-1} \otimes S\right), \tag{6.1}$$

where id again denotes the identity operation (*ideal* or *noiseless* channel).

The Schrödinger picture equivalent of this model was introduced by Bowen and Mancini in [BM04] and has been shown to encompass channels with Markovian correlated noise discussed previously in [MP02, MPV04, DWM03, BDB04]. As advertised in the Introduction, in Section 6.3 we will show that this model is sufficiently general to describe all causal quantum channels, which was left as an open problem in [BM04]. However, before we prove this structure theorem we will extend the notion of channel capacity from the memoryless setting to channels with memory, and we will present several different setups in which these channels may be operated for the transmission of both classical and quantum information.

6.2.2 Channel Capacity

As explained in Sec. 6.1.1, the standard definition of channel capacity applies also to quantum channels with memory. However, as illustrated in Fig. 6.3 we have to specify how to handle the initial and final memory states. In particular, we need to distinguish between setups in which Alice has control over the initial memory input state and may use it for the encoding procedure, and setups in which a malicious third party (Eve, say) controls the initial memory input, and by her choice of the input state $\mu \in \mathcal{B}_*(\mathcal{H}_M)$ will try to prevent Alice and Bob from communicating over the channel. Likewise, we may consider setups in which the final memory states are either ignored or accessible to Bob, and can thus be employed in the decoding process.

In the definition of channel capacity presented below, these four different scenarios are distinguished by a different range and domain of the encoding and decoding map, respectively, and give rise to four different channel capacities for both classical and quantum information transmission.

Definition 6.1. (Capacity for Memory Channels)
Let $\mathcal{H}_A$, $\mathcal{H}_B$, and $\mathcal{H}_M$ be Hilbert spaces. A positive number r is called an achievable rate *for the quantum memory channel $S\colon \mathcal{B}(\mathcal{H}_B)\otimes\mathcal{B}(\mathcal{H}_M) \to \mathcal{B}(\mathcal{H}_M)\otimes\mathcal{B}(\mathcal{H}_A)$ iff for any pair of integer sequences $(n_\nu)_{\nu\in\mathbb{N}}$ and $(m_\nu)_{\nu\in\mathbb{N}}$ with $\lim_{\nu\to\infty} n_\nu = \infty$ and $\overline{\lim}_{\nu\to\infty} \frac{m_\nu}{n_\nu} \leq r$ we have*

$$\lim_{\nu\to\infty} \Delta(n_\nu, m_\nu) = 0\,, \tag{6.2}$$

where we set

$$\Delta(n_\nu, m_\nu) := \inf_{E,D} \| E\, S_{n_\nu} D - \mathrm{id}_{\mathbb{C}^2}^{\otimes m_\nu} \|_{cb}\,, \tag{6.3}$$

the infimum taken over all encoding channels E and decoding channels D with suitable domain and range. The quantum channel capacity *$Q(S)$ of the memory channel S is defined to be the supremum of all achievable rates.*
In the different setups described above, the domain of the encoding channels E may or may not include the initial memory algebra $\mathcal{B}(\mathcal{H}_M)$, and the range of the decoding channels D may or may not contain the final memory algebra $\mathcal{B}(\mathcal{H}_M)$, resulting in four different quantum capacities $Q_{AB}(S)$, $Q_{AE}(S)$, $Q_{EB,\mu}(S)$, and $Q_{EE,\mu}(S)$, where the first index stands for the party (Alice, Bob, or Eve) who controls the initial memory state, the second index stands for the party who has access to the final memory state, and $\mu \in \mathcal{B}_*(\mathcal{H}_M)$ stands for Eve's choice of the initial memory state, if applicable.

Evidently, this definition is a straightforward generalization of the capacity of a memoryless channel, as presented in Def. 5.1, and reduces to this case if no memory system is present. A few remarks are in order:

Remark 6.2. The capacity of a quantum memory channel S for the transmission of *classical* information can be defined along the same lines, restricting encoding channels to *preparations* and decoding channels to *measurements*, and replacing the ideal qubit channel $\mathrm{id}_{\mathbb{C}^2}$ by the ideal bit channel in Eq. (6.3). The respective classical capacities are denoted by $C_{AB}(S)$, $C_{AE}(S)$, $C_{EB,\mu}(S)$, and $C_{EE,\mu}(S)$, and are no smaller than their quantum counterparts.

Remark 6.3. In the sections to follow, we will write $Q_*(S)$ and $C_*(S)$ whenever a certain statement holds for all the four channel capacities introduced in Def. 6.1, regardless of Eve's choice of the initial memory state.

Remark 6.4. It is obvious from the definition that for every memory channel S the capacities introduced in Def. 6.1 satisfy the following chain of inequalities:

$$Q_{EE,\mu}(S) \leq \{Q_{AE}(S), Q_{EB,\mu}(S)\} \leq Q_{AB}(S) \tag{6.4}$$

for all $\mu \in \mathcal{B}_*(\mathcal{H}_M)$, and accordingly for the classical capacities $C_{EE,\mu}(S)$ etc.

Remark 6.5. We have seen in Sec. 5.2 that there is considerable freedom in the definition of the channel capacity of a memoryless channel. Some of these variations are based solely on the properties of the error function $\Delta(n_\nu, m_\nu)$, and thus immediately apply to channels with memory as well. In particular, it is sufficient to find one pair of integer sequences $(n_\nu)_{\nu\in\mathbb{N}}$ and $(m_\nu)_{\nu\in\mathbb{N}}$ such that $\lim_{\nu\to\infty} \frac{m_\nu}{n_\nu} = r$ and $\lim_{\nu\to\infty} \Delta(n_\nu, m_\nu) = 0$, provided the diverging sequence $(n_\nu)_{\nu\in\mathbb{N}}$ is *subexponential* in the sense of Sec. 5.2.1, i. e., $\lim_{\nu\to\infty} \frac{n_{\nu+1}}{n_\nu} = 1$. In addition, the cb-norm in Eq. (6.3) can of course be replaced by equivalent distance measures such as *minimum fidelity* or *entanglement fidelity*, or even *average fidelity* and *channel fidelity.*

6.2.3 Examples

In the following, in order to illustrate the concepts introduced above we will present several examples of quantum memory channels. These will also serve to show that the various capacities introduced in Def. 6.1 may or may not coincide, thereby justifying our defining more than one capacity.

A simple model channel for which all the capacities introduced above coincide is the *shift channel* S^s. In principle, this is just a noiseless channel, but it interchanges memory and input register: $S^s(B\otimes M) = B\otimes M$ (Note that in the tensor representation that we have chosen, the identity channel id comes with the inherent flip, i. e., $\mathrm{id}\,(B\otimes M) = M\otimes B$.) Thus, in an n-fold concatenation of shift channels, the signals that Alice sends through the channel will be received by Bob undistorted one time step later. In the capacity limit of long messages (as $n\to\infty$), the initial qubit that Bob may lose if Eve controls the initial memory state, and the final qubit that he may lose if he cannot access the final memory state both have a negligible impact on the transmission rate, and therefore $Q_{EE,\mu}(S^s) = \lim_{n\to\infty} \frac{n-2}{n} \operatorname{ld} d = \operatorname{ld} d$ for all memory input states $\mu \in \mathcal{B}_*(\mathcal{H}_M)$, with $d := \dim \mathcal{H}_A = \dim \mathcal{H}_B = \dim \mathcal{H}_M$. Therefore, by Eq. (6.4) and Remark 6.2 all the above capacities equal $\operatorname{ld} d$. Further examples for channels in which the worst-case capacity and the best-case capacity are both maximal will be presented in Sec. 6.2.4.

An example of a memory channel in which the control over the initializing memory state can have a decisive influence on the channel performance is the channel with a *global classical switch*: Suppose that the memory algebra is a classical d-level system of diagonal $(d\times d)$ matrices, and that we are given a collection $\{T_i\}_{i=1}^d$ of d quantum memoryless channels $T_i\colon \mathcal{B}\to\mathcal{A}$. Then a quantum memory channel $S\colon \mathcal{B}\otimes\mathcal{M}\to\mathcal{M}\otimes\mathcal{A}$ with a global classical switch (d settings) is given by

$$S(B\otimes M) = \sum_{i=1}^{d} \langle i|M|i\rangle \; |i\rangle\langle i| \otimes T_i(B) \tag{6.5}$$

for all $B\in\mathcal{B}$ and $M\in\mathcal{M}$. In an n-fold concatenation of this channel, the channel T_i is applied in every time step if the initial memory input state was $|i\rangle\langle i|$. If Alice initially

sends a pre-defined sequence of test states, Bob may find out what the initial memory setting was and choose the decoding channel accordingly. Thus, the best case capacity in this setting will be $\max_{i=1,..d}\{Q(T_i)\}$, and the worst case capacity will be no larger than $\min_{i=1,..d}\{Q(T_i)\}$. These two may clearly differ.

6.2.4 Pure Channels

Pure memory channels are channels which have only one Kraus operator in Eq. (2.36). From the unitality condition, $S(\mathbb{1}) = \mathbb{1}$, it is then clear that these channels have a Kraus representation $S(B \otimes M) = V^*(B \otimes M)V$ with isometric $V : \mathcal{H}_M \otimes \mathcal{H}_A \to \mathcal{H}_B \otimes \mathcal{H}_M$.

In this Section we will show that for pure channels with finite memory, the various capacities introduced in Def. 6.1 coincide and are maximal, i. e., we have the following

Theorem 6.6. (Capacity of Pure Channels)
Let $S\colon \mathcal{B}(\mathcal{H}_B) \otimes \mathcal{B}(\mathcal{H}_M) \to \mathcal{B}(\mathcal{H}_M) \otimes \mathcal{B}(\mathcal{H}_A)$ be a pure quantum memory channel with finite-dimensional memory algebra $\mathcal{B}(\mathcal{H}_M)$. With the convention introduced in Remark 6.3 we then have:

$$Q_*(S) = \min\{\mathrm{ld}\,\dim\mathcal{H}_A, \mathrm{ld}\,\dim\mathcal{H}_B\} = C_*(S)\,. \tag{6.6}$$

Our strategy for the proof is to show that for pure channels it is possible to satisfy the Knill-Laflamme error correction criteria [KL97], which imply that perfect signal recovery can be achieved. This is more than what is required for capacity purposes, since the definition of channel capacity, as presented in Sec. 6.2.2, only demands that errors vanish asymptotically, i. e., in the limit of long messages $n \to \infty$.

Since we will have to refer to them repeatedly in the course of the proof, we start by restating the Knill-Laflamme conditions for perfect error correction (cf. Th. 10.1 in [NC00]): a necessary and sufficient condition for a quantum channel $T\colon \mathcal{B}(\mathcal{H}_2) \to \mathcal{B}(\mathcal{H}_1)$ with Kraus operators $\{t_i\}_{i=1}^K$ to be completely correctable on a subspace $\mathcal{K} \subset \mathcal{H}_1$ is the existence of an orthonormal basis $\{|\alpha\rangle\}_{\alpha=1}^{\dim\mathcal{K}}$ of $\mathcal{K}$ such that

$$\langle\alpha|t_i^* t_j|\beta\rangle = \omega_{i,j}\,\langle\alpha|\beta\rangle, \tag{6.7}$$

where the coefficients $\omega_{i,j} \in \mathbb{C}$ are not permitted to depend on the basis labels α, β. If the orthonormal basis $\{|\alpha\rangle\}_\alpha \subset \mathcal{H}_1$ has $N \in \mathbb{N}$ elements, we say that there exists a quantum code of dimension N.

Coming back to pure channels, we see that in the setup in which Alice controls the initial memory state and Bob can read out the final memory state there is only one (isometric) Kraus operator V, and thus it is straightforward to satisfy Eq. (6.7) and achieve rates of up to $\min\{\mathrm{ld}\,\dim\mathcal{H}_A, \mathrm{ld}\,\dim\mathcal{H}_B\}$.

By Eq. (6.4) and Remark 6.2, in order to complete the proof of Th. 6.6 it is therefore sufficient to show that $Q_{EE,\mu}(S) \geq \min\{\mathrm{ld}\,\dim\mathcal{H}_A, \mathrm{ld}\,\dim\mathcal{H}_B\}\ \forall\,\mu \in \mathcal{B}_*(\mathcal{H}_M)$. Again

we will show that it is possible to satisfy the error-correction conditions Eq. (6.7). However, in the worst-case scenario in which Eve chooses an arbitrary input state $\mu \in \mathcal{B}_*(\mathcal{H}_M)$ and Bob has no control over the final memory output the resulting channel is no longer pure, but can be given a Kraus representation with no more than d_M^2 Kraus operators, where $d_M := \dim \mathcal{H}_M$:

Lemma 6.7. *Assume that $\mathcal{H}_A$, $\mathcal{H}_B$, and $\mathcal{H}_M$ are finite-dimensional Hilbert spaces, and let $d_M := \dim \mathcal{H}_M$. Assume further that $S: \mathcal{B}(\mathcal{H}_B) \otimes \mathcal{B}(\mathcal{H}_M) \to \mathcal{B}(\mathcal{H}_M) \otimes \mathcal{B}(\mathcal{H}_A)$ is a pure quantum memory channel, i. e., $S(B \otimes M) = V^*(B \otimes M)V$ for some isometric $V: \mathcal{H}_M \otimes \mathcal{H}_A \to \mathcal{H}_B \otimes \mathcal{H}_M$. Let $\hat{S}_\mu: \mathcal{B}(\mathcal{H}_B) \to \mathcal{B}(\mathcal{H}_A)$ be the restriction of S to the B-system, with fixed initial memory state $\mu \in \mathcal{B}_*(\mathcal{H}_M)$. Then $\hat{S}_\mu$ can be given a Kraus representation with d_M^2 Kraus operators.*

Proof: Let $\{|\alpha\rangle\}_{\alpha=1}^{d_M}$ be the eigenbasis of $\mu \in \mathcal{B}_*(\mathcal{H}_M)$, and suppose that $\{|i\rangle\}_{i=1}^{d_A}$ and $\{|j'\rangle\}_{j'=1}^{d_B}$ are orthonormal bases for $\mathcal{H}_A$ and $\mathcal{H}_B$, respectively. The Kraus isometry $V: \mathcal{H}_M \otimes \mathcal{H}_A \to \mathcal{H}_B \otimes \mathcal{H}_M$ can then be given the representation

$$V = \sum_{\alpha,\beta=1}^{d_M} V_{\alpha,\beta} \otimes |\alpha\rangle\langle\beta| \tag{6.8}$$

with operators $V_{\alpha,\beta} = \sum_{i=1}^{d_A} \sum_{j'=1}^{d_B} \langle j', \alpha|V|\beta, i\rangle\, |j'\rangle\langle i|$. From Eq. (6.8) we see that for arbitrary $\varrho \in \mathcal{B}_*(\mathcal{H}_A)$ and $B \in \mathcal{B}(\mathcal{H}_B)$ we have

$$\begin{aligned} \operatorname{tr}\,(\varrho \otimes \mu)\, V^*(B \otimes \mathbb{1}_M)V &= \sum_{\alpha,\beta,\gamma=1}^{d_M} \operatorname{tr}(\varrho\, V_{\alpha,\gamma}^*\, B\, V_{\alpha,\beta})\, \langle\beta|\mu|\gamma\rangle \\ &= \sum_{\alpha,\beta=1}^{d_M} \mu_\beta \operatorname{tr}(\varrho\, V_{\alpha,\beta}^*\, B\, V_{\alpha,\beta}) \\ &= \sum_{\alpha,\beta=1}^{d_M} \operatorname{tr}\,\varrho\, \hat{s}_{\mu,\alpha\beta}^*\, B\, \hat{s}_{\mu,\alpha\beta} \\ &= \operatorname{tr}\,\varrho\, \hat{S}_\mu(B), \end{aligned} \tag{6.9}$$

where we have set $\hat{s}_{\mu,\alpha\beta} := \sqrt{\mu_\beta}\, V_{\alpha,\beta}$, and $\{\mu_\beta\}_{\beta=1}^{d_M}$ are the eigenvalues of $\mu \in \mathcal{B}_*(\mathcal{H}_M)$. Thus, the restricted channel $\hat{S}_\mu$ can be given a representation with d_M^2 Kraus operators, as claimed. ■

Note that the number of Kraus operators is independent of the dimension of both Alice's and Bob's systems $\mathcal{H}_A$ and $\mathcal{H}_B$, and thus the above result holds true also for the concatenated memory channel $S_n: \mathcal{B}(\mathcal{H}_B)^{\otimes n} \otimes \mathcal{B}(\mathcal{H}_M) \to \mathcal{B}(\mathcal{H}_M) \otimes \mathcal{B}(\mathcal{H}_A)^{\otimes n}$, independently of $n \in \mathbb{N}$. Consequently, in the limit $n \to \infty$ of long messages our setup corresponds to a channel with large input space interacting with a small environment. Physical intuition suggests that in such a setup the loss of information to the environment should be negligible, and it should be possible to operate the channel like an almost ideal one. This is the essence of the following

Lemma 6.8. *Let $T: \mathcal{B}(\mathcal{H}) \to \mathcal{B}(\mathcal{H})$ be a channel with K Kraus operators. Then there exists a quantum code of dimension at least $\left\lfloor \frac{\dim \mathcal{H}}{2^{K^2}} \right\rfloor$.*

Proof: Let $\{t_i\}_{i=1}^K$ be a set of Kraus operators for T, and let $\tau_{i,j} := t_i^* t_j$. In order to find a subspace $\mathcal{K} \subset \mathcal{H}$ of high dimensionality such that the Knill-Laflamme conditions Eq. (6.7) are satisfied, the following strategy may seem promising: choose a state vector $\varphi_1 \in \mathcal{H}$ arbitrarily, and then choose

$$\varphi_2 \in \mathcal{K}_1 := \varphi_1^\perp \cap \bigcap_{i,j=1}^{K} (\tau_{i,j}\varphi_1)^\perp . \tag{6.10}$$

Iterate this procedure of successive removal of dimensions until no further state vectors can be found. In every step, at most K^2 dimensions are removed, so this strategy yields a subspace of dimension $\geq \frac{\dim \mathcal{H}}{K^2}$. Unfortunately, this procedure does not guarantee that inner products $\langle \varphi_\alpha | \tau_{i,j} | \varphi_\beta \rangle$ are independent of the basis labels, as required by the Knill-Laflamme conditions Eq. (6.7). However, this can be accomplished by a carefully balanced pairing of eigenvectors, at the expense of a smaller code space.

Note that any operator $\tau \in \mathcal{B}(\mathcal{H})$ can be written as the weighted sum of two Hermitian operators, $\tau = \frac{1}{2}\tau_+ + \frac{i}{2}\tau_-$ with $\tau_+ := \tau^* + \tau$ and $\tau_- := i(\tau^* - \tau)$. Since the Knill-Laflamme conditions Eq. (6.7) are linear in the operators $\tau_{i,j}$, we may assume without loss that all operators $\tau_{i,j}$ are Hermitian. Let τ be one of these operators, and let $\{\lambda_\alpha\}_{\alpha=1}^d$ be the set of its eigenvalues, where $d := \dim \mathcal{H}$ and multiple eigenvalues appear according to their multiplicity. Choose $\omega \in \mathbb{R}$ such that equally many of the real numbers $\mu_\alpha := \lambda_\alpha - \omega$ lie on the positive and on the negative axis. (If necessary, reduce the dimension of $\mathcal{H}$ by one.) Now, if ψ_α is some eigenvector of the operator $\tau - \omega \mathbb{1}$ corresponding to the eigenvalue $\mu_\alpha > 0$, and $\psi_{-\alpha}$ is an eigenvector corresponding to the eigenvalue $\mu_{-\alpha} < 0$, by setting

$$\varphi_\alpha := \frac{1}{\sqrt{1 - \frac{\mu_\alpha}{\mu_{-\alpha}}}} \left(\psi_\alpha + \sqrt{\frac{\mu_\alpha}{-\mu_{-\alpha}}} \psi_{-\alpha} \right), \tag{6.11}$$

we obtain a Hilbert space $\mathcal{K}_1 := \operatorname{lin} \left\{ \varphi_\alpha \mid \alpha = 1, ..., \frac{d}{2} \right\}$ of dimension $\frac{d}{2}$ satisfying the Knill-Laflamme conditions Eq. (6.7) for the operator τ, i. e.,

$$\langle \varphi_\alpha | \tau - \omega \mathbb{1} | \varphi_\beta \rangle = 0 \qquad \forall \ \alpha, \beta = 1, ..., \frac{d}{2}. \tag{6.12}$$

Now, choose another operator $\tau' \in \{\tau_{i,j}\}_{i,j=1}^K$ and repeat the above pairing procedure on the subspace $\mathcal{K}_1$, resulting in a subspace $\mathcal{K}_2 \subset \mathcal{H}$ of dimension $\frac{d}{4}$. After K^2 steps, the resulting subspace has dimension at least $\frac{d}{2^{K^2}}$, which is the desired result. ■

We can now complete the proof of Th. 6.6: Applying the Knill-Laflamme code described in the proof of Lemma 6.8 to the concatenated memory channel $\hat{S}_{\mu,n}$ with d_M^2 Kraus operators, we immediately see that for all $\mu \in \mathcal{B}_*(\mathcal{H}_M)$

$$Q_{EE,\mu}(S) \geq \lim_{n \to \infty} \frac{1}{n} \operatorname{ld} \frac{d^n}{2^{d_M^4}} = \operatorname{ld} d \, , \tag{6.13}$$

where $d := \min\{\text{ld dim}\,\mathcal{H}_A, \text{ld dim}\,\mathcal{H}_B\}$, as claimed. ■

Closely related results on channels interacting with *small* environments have been obtained independently by G. Bowen and S. Mancini [Bow04^a]. These authors also show that for such channels the Knill-Laflamme error correction conditions can be fulfilled. However, instead of the pairing of eigenvalues described in the proof of Lemma 6.8, their approach uses convex set arguments of Knill et al. [KLV00], which are based on a generalization of Radon's theorem [Tve66]. Our approach seems more straightforward, but this comes at the expense of a weaker estimate, since the more sophisticated strategy of Knill *et al.* yields a code of dimension $\geq \frac{d}{K^2(K^2+1)}$.

6.3 The Structure of Causal Channels

In the first part of this Chapter we have followed a constructive approach to quantum channels with memory, in the sense that quantum channels which process long messages were always thought of as concatenations of smaller units which process one quantum signal each. In this Section we take the alternative view and assume that we are a priori given a quantum channel on a long (possibly infinite) message string. Our interest is then in the internal structure of such a quantum channel. As advertised in Sec. 6.1.1, we will show in Th. 6.10 that under very general assumptions it can be decomposed into a chain of quantum memory channels.

This structure theorem requires some mathematical background from the theory of infinite-dimensional quantum systems and channel representations, most notably quasi-local algebras and the uniqueness of the minimal Stinespring dilation. The relevant material on quasi-local algebras is collected in Appendix C.

To set the stage, imagine that we have at our disposal a quantum channel which, at every discrete time step, transforms an input state on some observable algebra $\mathcal{A}$ into an output state on some (possibly different) observable algebra $\mathcal{B}$. It is represented (in Heisenberg picture) by a completely positive and unital map $T\colon \mathcal{B}_{\mathbb{Z}} \to \mathcal{A}_{\mathbb{Z}}$ between the quasi-local algebras $\mathcal{A}_{\mathbb{Z}}$ and $\mathcal{B}_{\mathbb{Z}}$ on Alice's and Bob's side of the channel, respectively. In the following, we will restrict ourselves to translational invariant channels, i. e., we assume that T commutes with the shift on the spin chain: $\sigma_{\mathcal{A}} \circ T = T \circ \sigma_{\mathcal{B}}$. In addition, we impose the physically reasonable constraint that outputs up to some time t do not depend on inputs at times $t' > t$, leading to the following

Definition 6.9. (Causal Channel)
A causal channel $T\colon \mathcal{B}_{\mathbb{Z}} \to \mathcal{A}_{\mathbb{Z}}$ *is a completely positive and unital translational invariant map such that for every* $z \in \mathbb{Z}$

$$T(B_{(-\infty,z]} \otimes \mathbb{1}_{[z+1,\infty)}) = T(B_{(-\infty,z]}) \otimes \mathbb{1}_{[z+1,\infty)} \tag{6.14}$$

for all $B_{(-\infty,z]} \in \mathcal{B}_{(-\infty,z]}$.

Bearing in mind that T is translational invariant, we will henceforth set $z = 0$, and we will use the short-hands $\mathcal{A}_- := \mathcal{A}_{(-\infty,0]}$ and $\mathcal{A}_+ := \mathcal{A}_{[1,\infty)}$ to denote the left and right half chain, respectively. $\mathcal{B}_-$ and $\mathcal{B}_+$ are defined analogously.

It is obvious from the definition that a concatenated memory channel satisfies the causality property Eq. (6.14). In this Section we will prove the converse: every causal channel can be represented as a concatenated memory channel (cf. Fig. 6.2).

Theorem 6.10. (Structure Theorem)
Let $T\colon \mathcal{B}_{\mathbb{Z}} \to \mathcal{A}_{\mathbb{Z}}$ be a causal channel. Ignore its outputs on the left half chain $\mathcal{B}_-$. Then we may find a memory observable algebra $\mathcal{M}$ and an initializing channel *$R\colon \mathcal{M} \to \mathcal{A}_-$ such that $\forall\, n \in \mathbb{N}$*

$$T(\mathbb{1}_- \otimes B_n) = (R \otimes \mathrm{id}_{\mathcal{A}}^{\otimes n})\, S_n(B_n \otimes \mathbb{1}_{\mathcal{M}}) \tag{6.15}$$

for all $B_n \in \mathcal{B}_{[1,n]} \simeq \mathcal{B}^{\otimes n}$, where S_n is the n-fold concatenation of a memory channel $S\colon \mathcal{B} \otimes \mathcal{M} \to \mathcal{M} \otimes \mathcal{A}$, as defined in Eq. (6.1) above.

If Eq. (6.15) holds, we say that the memory channel $S\colon \mathcal{B} \otimes \mathcal{M} \to \mathcal{M} \otimes \mathcal{A}$ generates the causal channel $T\colon \mathcal{B}_{\mathbb{Z}} \to \mathcal{A}_{\mathbb{Z}}$. In the finite-dimensional setup, a corresponding theorem has been proved by Eggeling *et al.* [ESW01]. Here we generalize this result to channels on quasi-local algebras. As in the finite-dimensional setting, the uniqueness of the minimal Stinespring representation will play a crucial role.

Proof of Th. 6.10: Let $\mathcal{H}$ be the Hilbert space associated with the *universal representation* (cf. Sec. 2.6) of the left half chain $\mathcal{A}_-$. Note that in general $\mathcal{H}$ will not be separable. However, we have seen in Sec. 2.5 that separability is not required in Stinespring's theorem. Suppose that $(\mathcal{K}, \pi, V)$ is a minimal Stinespring dilation for $T\,|_{\mathcal{B}_-}$, i. e.,

$$T(B) = V^*\, \pi(B)\, V \quad \forall\, B \in \mathcal{B}_- \tag{6.16}$$

for some Stinespring isometry $V\colon \mathcal{H} \to \mathcal{K}$. In the sequel, we will make repeated use of the Hilbert space isomorphism $\mathcal{H} \simeq \mathcal{H} \otimes (\mathbb{C}^d)^{\otimes n}$ (cf. Ch. 3 of Kreyszig's text [Kre89]), where $\mathcal{A} = \mathcal{B}(\mathbb{C}^d)$ for some $d \in \mathbb{N}$. From Stinespring's representation Eq. (6.16) and the causality property Eq. (6.14), we may then conclude that

$$\begin{aligned} V^*\, \pi\big(B \otimes \mathbb{1}_{\mathcal{B}}^{\otimes n}\big)\, V &= T\big(B \otimes \mathbb{1}_{\mathcal{B}}^{\otimes n}\big) \\ &= T(B) \otimes \mathbb{1}_{\mathcal{A}}^{\otimes n} \\ &= \big(V^* \otimes \mathbb{1}_{\mathcal{A}}^{\otimes n}\big)\, \big(\pi(B) \otimes \mathbb{1}_{\mathcal{A}}^{\otimes n}\big)\, \big(V \otimes \mathbb{1}_{\mathcal{A}}^{\otimes n}\big) \end{aligned} \tag{6.17}$$

for all $B \in \mathcal{B}_-$. Since V is a minimal dilation for T, so is $V \otimes \mathbb{1}_{\mathcal{A}}^{\otimes n}$ for $T \otimes \mathbb{1}_{\mathcal{A}}^{\otimes n}$. We may then conclude from Th. 2.7 that there exists an isometry $W_n\colon \mathcal{K} \otimes (\mathbb{C}^d)^{\otimes n} \to \mathcal{K}$ defined by

$$W_n\big(\pi(B) \otimes \mathbb{1}_{\mathcal{A}}^{\otimes n}\big)\big(V \otimes \mathbb{1}_{\mathcal{A}}^{\otimes n}\big)\, \psi \otimes \psi_n \;:=\; \pi\big(B \otimes \mathbb{1}_{\mathcal{A}}^{\otimes n}\big)\, V\, \psi \otimes \psi_n \tag{6.18}$$

for all $B \in \mathcal{B}_-$, $\psi \in \mathcal{H}$ and $\psi_n \in \mathcal{A}^{\otimes n}$ such that

$$\pi\big(B \otimes \mathbb{1}_{\mathcal{B}}^{\otimes n}\big)\, W_n = W_n\, \big(\pi(B) \otimes \mathbb{1}_{\mathcal{A}}^{\otimes n}\big) \tag{6.19}$$

for all $B \in \mathcal{B}_-$, and

$$W_n\, \big(V \otimes \mathbb{1}_{\mathcal{A}}^{\otimes n}\big) = V. \tag{6.20}$$

We are now in a position to reconstruct the memory algebra: Let $\mathcal{M} := \pi'(\mathcal{B}_-)$, the *commutant* of the observable algebra $\mathcal{B}_-$, and let $S_n\colon \mathcal{B}^{\otimes n} \otimes \mathcal{M} \to \mathcal{B}(\mathcal{K}) \otimes \mathcal{B}\big((\mathbb{C}^d)^{\otimes n}\big)$ be defined by

$$S_n(B \otimes M) := W_n^*\, \pi(B)\, M\, W_n \tag{6.21}$$

for all $B \in \mathcal{B}_-$ and $M \in \mathcal{M}$. The memory initializing channel $R\colon \mathcal{M} \to \mathcal{A}_-$ is given by

$$R(M) := V^*\, M\, V \qquad \forall\ M \in \mathcal{M}. \tag{6.22}$$

In order to justify these choices, we will first show that

$$S_n(\mathcal{B}^{\otimes n} \otimes \mathcal{M}) \subset \mathcal{M} \otimes \mathcal{A}^{\otimes n}. \tag{6.23}$$

Noting that $\pi\big(\mathbb{1}_{\mathcal{B}_-} \otimes \mathcal{B}^{\otimes n}\big)\, \mathcal{M} \subset \pi'\big(\mathcal{B}_- \otimes \mathbb{1}_B^{\otimes n}\big)$, we see from Eq. (6.19) that

$$\begin{aligned} W_n^*\, \pi\big(\mathbb{1}_{\mathcal{B}_-} \otimes B_n\big)\, M\, W_n\, \Big(\pi\big(\tilde{B}_{\mathcal{B}_-}\big) \otimes \mathbb{1}_A^{\otimes n}\Big) \\ &= W_n^*\, \pi\big(\mathbb{1}_{\mathcal{B}_-} \otimes B_n\big)\, M\, \pi\big(\tilde{B}_{\mathcal{B}_-} \otimes \mathbb{1}_B^{\otimes n}\big)\, W_n \\ &= W_n^*\, \pi\big(\tilde{B}_{\mathcal{B}_-} \otimes \mathbb{1}_B^{\otimes n}\big)\, \pi\big(\mathbb{1}_{\mathcal{B}_-} \otimes B_n\big)\, M\, W_n \\ &= \Big(\pi\big(\tilde{B}_{\mathcal{B}_-}\big) \otimes \mathbb{1}_A^{\otimes n}\Big)\, W_n^*\, \pi\big(\mathbb{1}_{\mathcal{B}_-} \otimes B_n\big)\, M\, W_n \end{aligned} \tag{6.24}$$

for all $B_n \in \mathcal{B}^{\otimes n}$ and $\tilde{B}_{\mathcal{B}_-} \in \mathcal{B}_-$, implying that

$$\Big[S_n(B_n \otimes m)\ \big|\ \pi\big(\tilde{B}_{\mathcal{B}_-}\big) \otimes \mathbb{1}_{\mathcal{A}}^{\otimes n}\Big]\ =\ 0\,, \tag{6.25}$$

from which Eq. (6.23) directly follows. To complete the proof, it suffices to show that S_n has the right concatenation properties, i. e.,

$$R(M) \;=\; \big(R \otimes \mathrm{id}_{\mathcal{A}}^{\otimes n}\big)\, S_n\big(\mathbb{1}_{\mathcal{B}}^{\otimes n} \otimes M\big) \qquad \text{and} \tag{6.26}$$

$$T(B) \;=\; \big(R \otimes \mathrm{id}_{\mathcal{A}}^{\otimes n}\big)\, S_n\big(B \otimes \mathbb{1}_{\mathcal{M}}\big) \tag{6.27}$$

for all $M \in \mathcal{M}$ and $B \in \mathcal{B}^{\otimes n}$. However, this is immediate from the definitions of S_n and R and Eq. (6.20). The result then follows by setting $S := S_1$. ■

As can be seen from the above reasoning, the commutant algebra $\mathcal{M}$ can be replaced by the von Neumann algebra generated by all elements $(\mathrm{id}_{\mathcal{K}} \otimes \omega_n)\, S_n(B_n \otimes \mathbb{1}_{\mathcal{M}})$. In the above construction there is no unique way of choosing the memory algebra: given an infinite chain of memory channels with memory algebra $\mathcal{M}$, considering it as a causal channel and applying the memory reconstruction as in the proof of Th. 6.10 will in general yield a different memory algebra $\mathcal{M}' \neq \mathcal{M}$.

It is clear from the proof of Th. 6.10 that the channel reconstruction will in general explicitly depend on the input initializer R, which describes the influence of input states in the remote past on the memory. In the following Section we will turn our attention to an important class of memory channels for which the memory initializer becomes completely irrelevant. These so-called *forgetful channels* therefore bridge the axiomatic and the constructive approach to quantum channels with memory. We will also show that generic memory channels are forgetful.

6.4 Forgetful Quantum Channels

Forgetful channels are quantum memory channels $S\colon \mathcal{B}\otimes\mathcal{M}\to\mathcal{M}\otimes A$ in which the effect of the initializing memory state dies away with time. More formally, we have the following

Definition 6.11. (Forgetful Quantum Channel)
Let $S\colon \mathcal{B}\otimes\mathcal{M}\to\mathcal{M}\otimes\mathcal{A}$ be a quantum memory channel, S_n its n-fold concatenation, and let $\hat{S}_n\colon \mathcal{M}\to\mathcal{M}\otimes\mathcal{A}^{\otimes n}$ be the concatenated channel in which Bob's outputs are ignored: $\hat{S}_n(M) := S_n(\mathbb{1}_{\mathcal{B}}^{\otimes n}\otimes M)$ for all $M\in\mathcal{M}$. Then S is called forgetful *iff there exists a sequence of quantum channels $\tilde{S}_n\colon \mathcal{M}\to\mathcal{A}^{\otimes n}$ such that*

$$\lim_{n\to\infty}\|\hat{S}_n - \mathbb{1}_{\mathcal{M}}\otimes\tilde{S}_n\|_{cb} = 0\,. \tag{6.28}$$

We say that a causal channel $T\colon \mathcal{B}_{\mathbb{Z}}\to\mathcal{A}_{\mathbb{Z}}$ is forgetful *iff there exists a forgetful memory channel $S\colon \mathcal{B}\otimes\mathcal{M}\to\mathcal{M}\otimes\mathcal{A}$ that generates T.*

As a first illustrative example, let us briefly consider the classically mixed channel $S := p\,\mathrm{id} + (1-p)\,S^s$, where $p\in[0,1)$, and S^s denotes the shift channel introduced in Sec. 6.2.3. When this channel is concatenated, in every step either the ideal channel or the shift channel is chosen with probabilities p and $1-p$, respectively. The only possible way for an n-fold concatenation $\hat{S}_n$ not to be forgetful is to choose the ideal channel id in every step. However, the probability for this event is p^n, and thus vanishes in the limit $n\to\infty$, implying that Eq. (6.28) holds.

In this simple example, a forgetful quantum channel arises from mixing the non-forgetful channel id with an arbitrarily small fraction of the forgetful channel S^s. This already seems to suggest that the property of *not* being forgetful is quite sensitive to perturbations. In Sec. 6.4.4 we will turn this idea into a rigorous proof showing that generic quantum channels are indeed forgetful.

Remark 6.12. Note that Def. 6.11 can be relaxed by requiring only that $(\tilde{S}_n)_{n\in\mathbb{N}}$ is a sequence of linear maps, yet not necessarily channels. To see that this leads to an equivalent definition of forgetfulness, assume that $\|\hat{S}_n - \mathbb{1}_{\mathcal{M}}\otimes\tilde{S}_n\|_{cb}\leq\varepsilon$ for some $\varepsilon>0$, $n\in\mathbb{N}$, and some linear operator $\tilde{S}_n$. Replacing $\mathbb{1}_{\mathcal{M}}\otimes\tilde{S}_n$ with the quantum channel

$(P \otimes \mathrm{id}_{\mathcal{A}}^{\otimes n}) \circ \hat{S}_n$, where $P\colon \mathcal{M} \to \mathbb{C} \circ \mathbb{1}_{\mathcal{M}}$ is the completely depolarizing channel, we see that

$$\begin{aligned} \|\hat{S}_n - (P \otimes \mathrm{id}_{\mathcal{A}}^{\otimes n}) \circ \hat{S}_n\|_{cb} &\leq \|\hat{S}_n - \mathbb{1}_{\mathcal{M}} \otimes \tilde{S}_n\|_{cb} \\ &\quad + \|\big(P \otimes \mathrm{id}_{\mathcal{A}}^{\otimes n}\big) \circ \big(\mathbb{1}_{\mathcal{M}} \otimes \tilde{S}_n - \hat{S}_n\big)\|_{cb} \\ &\leq 2\,\|\hat{S}_n - \mathbb{1}_{\mathcal{M}} \otimes \tilde{S}_n\|_{cb} \\ &\leq 2\,\varepsilon\,, \end{aligned} \tag{6.29}$$

and thus $\lim_{n\to\infty} \|\hat{S}_n - (P \otimes \mathrm{id}_{\mathcal{A}}^{\otimes n}) \circ \hat{S}_n\|_{cb} = 0$, implying that S is indeed forgetful in the sense of Def. 6.11.

6.4.1 Forgetfulness Criteria

There exist several equivalent criteria for a quantum memory channel to be forgetful. In particular, it is sufficient to show that the norm distance $\|\hat{S}_n - \mathbb{1}_{\mathcal{M}} \otimes \tilde{S}_n\|_{cb}$ falls below 1 for *some* $n \in \mathbb{N}$. What is more important, the memory effects can always be assumed to vanish exponentially fast. In addition, if the memory algebra $\mathcal{M}$ has finite dimension, the cb-norm criterion Eq. (6.28) can be replaced by the usual operator norm $\|\cdot\|_\infty$. In fact, we have the following

Proposition 6.13. (Forgetfulness)
Let $S\colon \mathcal{B} \otimes \mathcal{M} \to \mathcal{M} \otimes \mathcal{A}$ be a quantum memory channel, and for $n \in \mathbb{N}$ let $\hat{S}_n$ be defined as in Def. 6.11 above. Then S is forgetful iff there exists an integer $N \in \mathbb{N}$ and some linear operator $\tilde{S}_N\colon \mathcal{M} \to \mathcal{A}^{\otimes N}$ (not necessarily a channel) such that

$$\|\hat{S}_N - \mathbb{1}_{\mathcal{M}} \otimes \tilde{S}_N\|_{cb} < 1\,. \tag{6.30}$$

Assume in addition that the memory algebra $\mathcal{M}$ has finite dimension. Then S is forgetful iff for every $M \in \mathcal{M}$ and $\varepsilon > 0$ we may find a positive integer $N \in \mathbb{N}$ and $A_N \in \mathcal{A}^{\otimes N}$ such that

$$\|\hat{S}_N(M) - \mathbb{1}_{\mathcal{M}} \otimes A_N\|_\infty \leq \varepsilon\,\|M\|_\infty\,. \tag{6.31}$$

As advertised above, in the proof of Prop. 6.13 we will also be concerned with the speed of convergence in Eq. (6.28). In this context, the following simple Lemma will be helpful:

Lemma 6.14. *Let $(d_n)_{n\in\mathbb{N}}$ be a positive and non-increasing sequence satisfying the subadditivity inequality*

$$d_{n+m} \leq d_n\, d_m \quad \forall\, n, m \in \mathbb{N}. \tag{6.32}$$

Assume further that $d_N < 1$ for some $N \in \mathbb{N}$. Then

$$d_n \leq c^n \quad \forall\, n \geq N \tag{6.33}$$

for some constant $c < 1$, i. e., $(d_n)_{n\in\mathbb{N}}$ vanishes exponentially.

Proof of Lemma 6.14: Assume that $d_N < 1$ for some $N \in \mathbb{N}$. From the subadditivity inequality (6.32) we then see that $d_{N+N} \leq d_N^2$, and, by induction, $d_{\nu N} \leq d_N^\nu$ for all integers $\nu \in \mathbb{N}$. By the monotonicity of $(d_n)_{n\in\mathbb{N}}$ we may then conclude that for $n \in [\nu N, (\nu+1)N)$ we have

$$d_n \leq d_{\nu N} \leq d_N^\nu \leq \left(d_N^{\frac{1}{2N}}\right)^n = c^n \tag{6.34}$$

with $c := d_N^{\frac{1}{2N}} < 1$, as advertised. ■

For the second part of the proof of Prop. 6.13, we obviously need to bound the cb-norm $\|\cdot\|_{cb}$ of a linear operator $R\colon \mathcal{B}(\mathcal{H}_M) \to \mathcal{A}$ with $\dim \mathcal{H}_M < \infty$ in terms of its operator norm $\|\cdot\|_\infty$. This is the essence of the following

Lemma 6.15. *Let $R\colon \mathcal{B}(\mathcal{H}_M) \to \mathcal{A}$ be a linear operator, with $d_M := \dim \mathcal{H}_M < \infty$. We then have*

$$\|R\|_{cb} \leq d_M^2 \, \|R\|_\infty \, . \tag{6.35}$$

Proof of Lemma 6.15: By definition we have $\|R\|_{cb} = \sup_k\{\|R \otimes \mathrm{id}_k\|_\infty\}$, where id_k is the identity operation on the $(k \times k)$ matrices $\mathcal{M}_k$. Every $X \in \mathcal{B}(\mathcal{H}_M) \otimes \mathcal{M}_k$ can be given the expansion

$$X = \sum_\alpha M_\alpha \otimes K_\alpha = \sum_\alpha \sum_{i,j=1}^{d_M} \mu_{\alpha,ij} \, |i\rangle\langle j| \otimes K_\alpha = \sum_{i,j=1}^{d_M} |i\rangle\langle j| \otimes X_{ij} \, , \tag{6.36}$$

where we have set $X_{ij} := \sum_\alpha \mu_{\alpha,ij} \, K_\alpha$. Note that $\|X_{ij}\|_\infty \leq \|X\|_\infty \; \forall \; i,j = 1, ..., d_M$, implying that

$$\begin{aligned} \|\big(R \otimes \mathrm{id}_k\big)X\|_\infty &= \| \sum_{i,j=1}^{d_M} R(|i\rangle\langle j|) \otimes X_{ij}\|_\infty \\ &\leq \sum_{i,j=1}^{d_M} \|R\|_\infty \, \||i\rangle\langle j|\|_\infty \, \|X_{i,j}\|_\infty \\ &\leq d_M^2 \, \|R\|_\infty \, \|X\|_\infty \end{aligned} \tag{6.37}$$

holds independently of k. We thus have $\|R\|_{cb} = \sup_k\{\|R \otimes \mathrm{id}_k\|_\infty\} \leq d_M^2 \, \|R\|_\infty$, as claimed. ■

We now have the necessary tools at hand to tackle the

Proof of Prop. 6.13: We will first prove the first part of Prop. 6.13. Thus, at this point we make no assumptions on the dimensionality of $\mathcal{M}$. If S is forgetful, Eq. (6.30) is immediate from the definition. In order to prove the converse, let

$$d_n := \inf\left\{\|\hat{S}_n - \mathbb{1}_{\mathcal{M}} \otimes \tilde{S}_n\|_{cb} \;\middle|\; \tilde{S}_n\colon \mathcal{M} \to \mathcal{A}^{\otimes n}, \text{linear}\right\}. \tag{6.38}$$

for $n \in \mathbb{N}$. Our strategy is to show that $(d_n)_{n\in\mathbb{N}}$ satisfies the conditions of Lemma 6.14. From Eq. (6.30) we can then conclude that $d_n \leq c^n$ for all $n \geq N$ for some constant $c < 1$, and thus S is forgetful with exponentially vanishing errors by Remark 6.12.

We start by showing that $(d_n)_{n\in\mathbb{N}}$ is non-increasing, i. e., $d_{n+1} \leq d_n \; \forall \; n \in \mathbb{N}$. From the definition of $\hat{S}_n$, we have

$$\begin{aligned} \hat{S}_{n+1} &= \big(\hat{S} \otimes \mathrm{id}_{\mathcal{A}}^{\otimes n}\big) \circ \hat{S}_n \\ &= \big(\hat{S} \otimes \mathrm{id}_{\mathcal{A}}^{\otimes n}\big) \circ \big(\hat{S}_n - \mathbb{1}_{\mathcal{M}} \otimes \tilde{S}_n\big) + \big(\hat{S} \otimes \mathrm{id}_{\mathcal{A}}^{\otimes n}\big) \circ \big(\mathbb{1}_{\mathcal{M}} \otimes \tilde{S}_n\big) \\ &= \big(\hat{S} \otimes \mathrm{id}_{\mathcal{A}}^{\otimes n}\big)\big(\hat{S}_n - \mathbb{1}_{\mathcal{M}} \otimes \tilde{S}_n\big) + \mathbb{1}_{\mathcal{M}} \otimes \mathbb{1}_{\mathcal{A}} \otimes \tilde{S}_n \,, \end{aligned} \tag{6.39}$$

where in the last step we have applied the unitality of $\hat{S}$. From Eq. (6.39) and unitality of the cb-norm we may conclude that

$$\begin{aligned} d_{n+1} &\leq \|\hat{S}_{n+1} - \mathbb{1}_{\mathcal{M}} \otimes \mathbb{1}_{\mathcal{A}} \otimes \tilde{S}_n\|_{cb} \\ &\leq \|\hat{S} \otimes \mathrm{id}_{\mathcal{A}}^{\otimes n}\|_{cb} \, \|\hat{S}_n - \mathbb{1}_{\mathcal{M}} \otimes \tilde{S}_n\|_{cb} \leq d_n \,, \end{aligned} \tag{6.40}$$

just as claimed. We will now show that $d_{n+m} \leq d_n \, d_m$ for all $n, m \in \mathbb{N}$. Similar to the above estimate, we have

$$\begin{aligned} \hat{S}_{n+m} &= \big(\hat{S}_n \otimes \mathrm{id}_{\mathcal{A}}^{\otimes m}\big) \, \hat{S}_m \\ &= \big(\hat{S}_n \otimes \mathrm{id}_{\mathcal{A}}^{\otimes m}\big) \, \big(\hat{S}_m - \mathbb{1}_{\mathcal{M}} \otimes \tilde{S}_m\big) + \big(\hat{S}_n \otimes \mathrm{id}_{\mathcal{A}}^{\otimes m}\big) \, \big(\mathbb{1}_{\mathcal{M}} \otimes \tilde{S}_m\big) \\ &= \Big[\big(\hat{S}_n - \mathbb{1}_{\mathcal{M}} \otimes \tilde{S}_n\big) \otimes \mathrm{id}_{\mathcal{A}}^{\otimes m}\Big] \, \big(\hat{S}_m - \mathbb{1}_{\mathcal{M}} \otimes \tilde{S}_m\big) + \mathbb{1}_{\mathcal{M}} \otimes \tilde{S}_{n+m} \,, \end{aligned} \tag{6.41}$$

where we have introduced the short hand

$$\tilde{S}_{n+m} := \mathbb{1}_{\mathcal{A}}^{\otimes n} \otimes \tilde{S}_m + \big(\tilde{S}_n \otimes \mathrm{id}_{\mathcal{A}}^{\otimes m}\big)\big(\hat{S}_m - \mathbb{1}_{\mathcal{M}} \otimes \tilde{S}_m\big) \,. \tag{6.42}$$

Invoking again both the unitality and the multiplicativity of the cb-norm, we may conclude from Eq. (6.41) that

$$\|\hat{S}_{n+m} - \mathbb{1}_{\mathcal{M}} \otimes \tilde{S}_{n+m}\|_{cb} \leq \|\hat{S}_n - \mathbb{1}_{\mathcal{M}} \otimes \tilde{S}_n\|_{cb} \, \|\hat{S}_m - \mathbb{1}_{\mathcal{M}} \otimes \tilde{S}_m\|_{cb} \leq d_n \, d_m \,, \tag{6.43}$$

which is the desired estimate. Note that $\tilde{S}_{n+m}$ is clearly linear and unital, but not necessarily positive. This is why we did not require the maps $\tilde{S}_n$ to be channels in the definition of the sequence $(d_n)_{n\in\mathbb{N}}$. This completes the first part of the proof. ▲

For the second part, assume that $\mathcal{M} = \mathcal{B}(\mathcal{H}_M)$ with $d_M := \dim \mathcal{H}_M < \infty$. If Eq. (6.31) holds, by the same reasoning as in Remark 6.12 we may conclude that $\mathbb{1}_{\mathcal{M}} \otimes A_N$ may be replaced by $(P \otimes \mathrm{id}_{\mathcal{A}}^{\otimes N}) \circ \hat{S}_N(M)$, implying that for every $M \in \mathcal{M}$ and $\varepsilon > 0$ we may find a positive integer $N \in \mathbb{N}$ such that

$$\|\hat{S}_N(M) - \big(P \otimes \mathrm{id}_{\mathcal{A}}^{\otimes N}\big) \circ \hat{S}_N(M)\|_\infty \leq 2\,\varepsilon \, \|M\|_\infty \,. \tag{6.44}$$

In order to arrive at a uniform bound, let us introduce an orthonormal basis $\{|i\rangle\}_{i=1}^{d_M}$ for $\mathcal{H}_M$. Since $\mathcal{H}_M$ has finite dimension, Eq. (6.44) holds uniformly for the basis operators

$\{|i\rangle\langle j|\}_{i,j=1}^{d_M}$ for some possibly larger N. Thus, by setting $M = \sum_{i,j=1}^{d_M} M_{i,j}\,|i\rangle\langle j|$ we see that

$$\begin{aligned}
&\|\hat{S}_N(M) - \big(P \otimes \mathrm{id}_{\mathcal{A}}^{\otimes N}\big) \circ \hat{S}_N(M)\|_\infty \\
&\qquad \leq \sum_{i,j=1}^{d_M} |M_{i,j}|\, \|\hat{S}_N(|i\rangle\langle j|) - \big(P \otimes \mathrm{id}_{\mathcal{A}}^{\otimes N}\big) \circ \hat{S}_N(|i\rangle\langle j|)\|_\infty \\
&\qquad \leq 2\,\varepsilon \sum_{i,j=1}^{d_M} |M_{i,j}| \\
&\qquad \leq\ 2\,\varepsilon\, d_M^2\, \|M\|_\infty\,,
\end{aligned} \tag{6.45}$$

where in the last step we have used that $|M_{i,j}| \leq \|M\|_\infty$ for all $i, j = 1, ..., d_M$. Making use of Lemma 6.15, we may conclude from Eq. (6.45) that

$$\|\hat{S}_N - \big(P \otimes \mathrm{id}_{\mathcal{A}}^{\otimes N}\big) \circ \hat{S}_N\|_{cb} \leq 2\,\varepsilon\, d_M^4\,. \tag{6.46}$$

Thus, choosing $\varepsilon < \frac{1}{2d_M^4}$, we may find an integer $N \in \mathbb{N}$ such that Eq. (6.30) holds. Therefore, S is forgetful by the first part of the proof. The converse is immediate from the definition of forgetfulness. ■

From the proof of Prop. 6.13 we may immediately deduce the following

Corollary 6.16. (Exponential Convergence)
Let $S\colon \mathcal{B} \otimes \mathcal{M} \to \mathcal{M} \otimes \mathcal{A}$ be a forgetful quantum channel. Then the effect of the initial memory vanishes exponentially fast, i. e., we may find a constant $c < 1$ such that

$$\|\hat{S}_n - (P \otimes \mathrm{id}_{\mathcal{A}}^{\otimes n}) \circ \hat{S}_n\|_{cb} < c^n \tag{6.47}$$

for all sufficiently large n.

For convenience, and because we will use it later in Sec. 6.6, in the following Proposition we show how the definition of forgetfulness translates into the Schrödinger picture language.

Proposition 6.17. (Forgetfulness in Schrödinger Picture)
Let $S\colon \mathcal{B} \otimes \mathcal{M} \to \mathcal{M} \otimes \mathcal{A}$ be a quantum channel. Let $\varepsilon > 0$, and for $n \in \mathbb{N}$ let $\hat{S}_n$ be defined as in Def. 6.11. Assume that

$$\|\hat{S}_n - \big(P \otimes \mathrm{id}_{\mathcal{A}}^{\otimes n}\big)\hat{S}_n\|_\infty \leq \varepsilon\,, \tag{6.48}$$

where $P\colon \mathcal{M} \to \mathbb{C}\,\mathbb{1}_{\mathcal{M}}$ is a completely depolarizing channel. We then have

$$\|\mathrm{tr}_{\mathcal{B}^{\otimes n}}\, S_{n*}\big(\varrho_1 - \varrho_2\big)\|_1 \leq 2\,\varepsilon \tag{6.49}$$

for all density operators $\varrho_1, \varrho_2 \in \mathcal{M}_* \otimes \mathcal{A}_*^{\otimes n}$ such that $\mathrm{tr}_{\mathcal{M}}\varrho_1 = \mathrm{tr}_{\mathcal{M}}\varrho_2$. Conversely, suppose that Eq. (6.49) holds. Then Eq. (6.48) holds with the substitution $\varepsilon \mapsto 2\,\varepsilon$.

In particular, if the quantum channel S is forgetful, then from Remark 6.12 we know that the condition in Eq. (6.48) is satisfied, and thus Eq. (6.49) holds. If in addition the memory algebra $\mathcal{M}$ is finite-dimensional, Eq. (6.48) is a necessary and sufficient criterion for forgetfulness by Prop. 6.13. By the above Proposition, Eq. (6.49) then gives a necessary and sufficient criterion for forgetfulness in the Schrödinger picture language.

Proof of Prop. 6.17: We recall from Sec. 2.7.2 that for any linear operator $T: \mathcal{B} \to \mathcal{A}$, the operator norm $\|T\|_\infty$ equals the norm of the adjoint operator on the dual space,

$$\|T\|_\infty = \sup_{\|\varrho\|_1 \leq 1} \|T_*(\varrho)\|_1 \,. \tag{6.50}$$

Now suppose that Eq. (6.48) holds. Since $\mathrm{id}_{\mathcal{A}_*}^{\otimes n} \otimes P_* = \mathrm{tr}_{\mathcal{M}}$, the partial trace on the memory algebra $\mathcal{M}$, we may conclude from Eq. (6.48) and the norm duality Eq. (6.50) that

$$\|\hat{S}_{n*}(\varrho) - \hat{S}_{n*}\,\mathrm{tr}_{\mathcal{M}}\varrho\|_1 \leq \varepsilon \qquad \forall\, \varrho \in \mathcal{M}_* \otimes \mathcal{A}_*^{\otimes n}, \tag{6.51}$$

which implies that for arbitrary $\varrho_1, \varrho_2 \in \mathcal{M}_* \otimes \mathcal{A}_*^{\otimes n}$ such that $\mathrm{tr}_{\mathcal{M}}\varrho_1 = \mathrm{tr}_{\mathcal{M}}\varrho_2$ we have

$$\|\hat{S}_{n*}(\varrho_1) - \hat{S}_{n*}(\varrho_2)\|_1 \leq 2\,\varepsilon \tag{6.52}$$

by the triangle inequality. Eq. (6.49) then follows by noting that $\hat{S}_{n*} = \mathrm{tr}_{\mathcal{B}^{\otimes n}} \circ S_{n*}$.

Conversely, from Eq. (6.49) we can conclude that

$$\|\hat{S}_{n*}\big(\varrho - \mathrm{tr}_{\mathcal{M}}\varrho\big)\|_1 \leq 2\,\varepsilon \qquad \forall\, \varrho \in \mathcal{M}_* \otimes \mathcal{A}_*^{\otimes n}\,, \tag{6.53}$$

which implies Eq. (6.48) (with the substitution $\varepsilon \mapsto 2\,\varepsilon$) by means of the norm duality Eq. (6.50). ■

6.4.2 Example: the Partial Flip Channel

Prop. 6.13 (and its Schrödinger dual Prop. 6.17) can be employed to test whether a given quantum memory channel is forgetful. As an illustrating example, let us consider the unitary *partial flip* operation

$$U_\eta := \cos\eta\,\mathbb{F} \,+\, i\sin\eta\,\mathbb{1} \tag{6.54}$$

with $\eta \in [0, 2\pi)$. As in Sec. 3.4, $\mathbb{F} := \sum_{i,j} |i\,j\rangle\langle j\,i|$ denotes the *flip operator*. Since $\mathbb{F}(B \otimes M)\mathbb{F} = M \otimes B$, for $\eta = 0$ the partial flip is just the *shift channel* S^s introduced in Sec. 6.2.3, which we already know is forgetful. With the help of Prop. 6.13, we will show that the partial flip is forgetful whenever $\cos\eta > \frac{7}{8}$. In fact, it is sufficient to prove that

$$\|U_\eta - \mathbb{F}\|_\infty < \frac{1}{2} \tag{6.55}$$

holds in the designated parameter range, since this will immediately imply that

$$\|U_\eta^* \, \mathbb{1}_\mathcal{B} \otimes (\cdot) \, U_\eta - \mathbb{F} \, \mathbb{1}_\mathcal{B} \otimes (\cdot) \, \mathbb{F}\|_{cb} \, < \, 1, \tag{6.56}$$

from which forgetfulness of the partial flip follows by Prop. 6.13. To see that Eq. (6.55) holds, set $\Delta_\eta := U_\eta - \mathbb{F}$ and observe that

$$\|\Delta_\eta^* \, \Delta_\eta\|_\infty \, = \, 2\,(1 - \cos\eta) \, < \, \frac{1}{4} \quad \Longleftrightarrow \quad \cos\eta > \frac{7}{8}. \tag{6.57}$$

We conjecture that the partial flip is in fact forgetful over the whole parameter range, apart from the trivial exceptions $\eta = \frac{1}{2}\pi$ and $\eta = \frac{3}{2}\pi$. Yet this forgetfulness is not witnessed by the criteria developed in Sec. 6.4.1.

6.4.3 Obedient Quantum Channels

The partial flip channel discussed in Sec. 6.4.2 has another interesting property. As pointed out by Ziman *et al.* [ZSB$^+$02, SZS$^+$02], in the asymptotic limit $n \to \infty$ it allows to prepare any desired memory output state by choosing some suitable sequence of input register states. We call a memory channel with this property *obedient*.

Definition 6.18. (Obedience)
Let $S_*\colon \mathcal{B}_*(\mathcal{H}_M) \otimes \mathcal{B}_*(\mathcal{H}_A) \to \mathcal{B}_*(\mathcal{H}_B) \otimes \mathcal{B}_*(\mathcal{H}_M)$ be a quantum memory channel, and S_{n*} its n-fold concatenation. Let

$$\mathrm{Im}_\mathcal{M}(S) := \big\{\mu \in \mathcal{B}_*(\mathcal{H}_M) \;\big|\; \exists\, \omega \in \mathcal{B}_*(\mathcal{H}_M) \otimes \mathcal{B}_*(\mathcal{H}_A) : \mathrm{tr}_\mathcal{B} \circ S_*(\omega) = \mu\big\} \tag{6.58}$$

denote the set of accessible memory output states. Then S is called obedient *iff for any state $\mu \in \mathrm{Im}_\mathcal{M}(S)$ there exists a sequence of register input states $\{\varrho_n\}_{n\in\mathbb{N}}$ with $\varrho_n \in \mathcal{B}_*(\mathcal{H}_A)^{\otimes n}$ for all $n \in \mathbb{N}$ such that*

$$\lim_{n\to\infty} \|\mathrm{tr}_{\mathcal{B}^{\otimes n}} \circ S_{n*}(\sigma \otimes \varrho_n) - \mu\|_1 = 0 \tag{6.59}$$

holds regardless of the memory input state $\sigma \in \mathcal{B}_*(\mathcal{H}_M)$.

Hence, for any desired final memory state $\mu \in \mathcal{B}_*(\mathcal{H}_M)$ in the range of S we may find a sequence $\{\varrho_n\}_{n\in\mathbb{N}}$ that will drive the memory output towards an approximation of μ. This can be of great practical relevance whenever the memory system itself is not directly accessible, but might still be steered by means of the register-memory interaction.

In the case of the partial flip channel, Ziman *et al.* [ZSB$^+$02, SZS$^+$02] have shown that the register input $\varrho_n = \mu^{\otimes n}$ results in an approximation of μ, with exponentially vanishing errors. This holds true over the entire parameter range, apart from the trivial exceptions $\eta = \frac{1}{2}\pi$ and $\eta = \frac{3}{2}\pi$. If the register inputs are interpreted as a spin bath in some thermal state ϱ, the partial flip channel thus describes a thermalization dynamics for the memory system. In fact, all forgetful quantum channels which are onto can be applied to prepare arbitrary memory output states:

Proposition 6.19. (Forgetfulness $\Longrightarrow$ Obedience)
Any forgetful memory channel $S_*\colon \mathcal{B}_*(\mathcal{H}_M) \otimes \mathcal{B}_*(\mathcal{H}_A) \to \mathcal{B}_*(\mathcal{H}_B) \otimes \mathcal{B}_*(\mathcal{H}_M)$ *such that* $\mathrm{Im}_{\mathcal{M}}(S) = \mathcal{B}_*(\mathcal{H}_M)$ *is obedient in the sense of Def. 6.18.*

Proof: This is straightforward. Let $\mu \in \mathcal{B}_*(\mathcal{H}_M)$ be the desired memory output state. Since $\mathrm{Im}_{\mathcal{M}}(S) = \mathcal{B}_*(\mathcal{H}_M)$, for any positive integer $n \in \mathbb{N}$ there exists a quantum state $\omega_n \in \mathcal{B}_*(\mathcal{H}_M) \otimes \mathcal{B}_*(\mathcal{H}_A)^{\otimes n}$ such that

$$\mathrm{tr}_{\mathcal{B}^{\otimes n}} \circ S_{n*}(\omega_n) = \mu\,. \tag{6.60}$$

Setting $\varrho_n := \mathrm{tr}_{\mathcal{M}}\omega_n$, we then immediately conclude from Prop. 6.17 and Eq. (6.60) that for any memory input state $\sigma \in \mathcal{B}_*(\mathcal{H}_M)$

$$\|\mathrm{tr}_{\mathcal{B}^{\otimes n}} S_{n*}(\sigma \otimes \varrho_n) - \mu\|_1 = \|\mathrm{tr}_{\mathcal{B}^{\otimes n}} S_{n*}(\sigma \otimes \varrho_n) - \mathrm{tr}_{\mathcal{B}^{\otimes n}} S_{n*}(\omega_n)\|_1 \leq \varepsilon \tag{6.61}$$

for all sufficiently large $n \in \mathbb{N}$. Hence, S is obedient according to Def. 6.18. ■

The converse of Prop. 6.19 is in general false: there are obedient quantum channels which are not forgetful. A simple example is a channel with a classical input control: Assume that $S_{i*}\colon \mathcal{B}_*(\mathcal{H}_M) \otimes \mathcal{B}_*(\mathcal{H}_A) \to \mathcal{B}_*(\mathcal{H}_B) \otimes \mathcal{B}_*(\mathcal{H}_M)$ are two quantum channels such that S_1 is forgetful (and hence obedient by Prop. 6.19), whereas S_2 is chosen non-forgetful. Appending a classical two-level input flag system $\mathcal{C}_2$, we define the quantum channel $S_*\colon \mathcal{C}_2^* \otimes \mathcal{B}_*(\mathcal{H}_M) \otimes \mathcal{B}_*(\mathcal{H}_A) \to \mathcal{B}_*(\mathcal{H}_B) \otimes \mathcal{B}_*(\mathcal{H}_M) \otimes \mathcal{C}_2^*$ by setting

$$S_*(\gamma \otimes \varrho) := \sum_{i=1}^{2} \langle i|\gamma|i\rangle\, S_{i*}(\varrho) \otimes |i\rangle\langle i| \tag{6.62}$$

for all $\varrho \in \mathcal{B}_*(\mathcal{H}_A)$ and $\gamma \in \mathcal{C}_2^*$. This channel allows to prepare arbitrary memory output states by choosing $\gamma = |1\rangle\langle 1|$ and suitable register input sequences, since S_1 is obedient. However, the channel is not forgetful, since non-vanishing memory effects may arise when Alice sets $\gamma = |2\rangle\langle 2|$. This establishes the counterexample.

In the special case of unitary quantum channels, the asymptotic preparation of memory output states has been investigated by Wellens *et al.* [WBK$^+$00] under the name *asymptotic completeness*. While asymptotic completeness and obedience are certainly related concepts, they seem to differ in fine points, for instance in the choice of the operator topology. Asymptotic completeness of the Jaynes-Cummings interaction, which governs the physics of the micromaser experiment described in Sec. 6.1.2, is claimed as a main mathematical result in [WBK$^+$00]. Unfortunately, a proof is not available in the cited literature [KM00].

6.4.4 Generic Forgetfulness

As advertised in Sec. 6.1.1, we will now establish that most memory channels are forgetful. Hence, the additional complications that occur due to the freedom in the memory initialization do not usually appear.

We will prove below that forgetful quantum channels are *dense* in the set of quantum memory channels: for every non-forgetful quantum channel we may find a forgetful memory channel which differs arbitrarily little from it. Thus, even the partial flip at $\eta = \frac{1}{2}\pi$ and $\eta = \frac{3}{2}\pi$ (i. e., the identity id) can be approximated by a forgetful quantum channel, though not necessarily a unitary one.

What is more, along the lines of the simple example presented after Def. 6.11 above, Prop. 6.13 can be applied to show that all quantum channels in a finite-size neighborhood of a given forgetful quantum channel are likewise forgetful, i. e., the set of forgetful quantum channels is *open*. Combined with the *denseness* of forgetful quantum channels, this justifies the claim that generic quantum memory channels are indeed forgetful.

Theorem 6.20. (Generic Forgetfulness)
The set of forgetful quantum channels is open and dense in the set of quantum memory channels in $\|\cdot\|_{cb}$-norm topology.

Proof: We will first show that the set of forgetful quantum channels is dense in the set of quantum memory channels. From any given (not necessarily forgetful) memory channel $S\colon \mathcal{B}\otimes\mathcal{M}\to\mathcal{M}\otimes\mathcal{A}$ we can easily construct a forgetful channel by mixing it with the completely depolarizing channel

$$D(B\otimes M) := \mathrm{tr}((B\otimes M)\delta)\ \mathbb{1}_{\mathcal{M}\otimes\mathcal{A}}\,, \tag{6.63}$$

where $\delta\in\mathcal{B}_*\otimes\mathcal{M}_*$ is an arbitrary quantum state. Just as in the classically mixed shift channel discussed above, all the terms in an n-fold concatenation of the mixed channel $S^\varepsilon := (1-\varepsilon)S + \varepsilon D$ yield the identity operator $\mathbb{1}_{\mathcal{M}}$ in the memory input, possibly apart from the S_n-contribution, which scales as $(1-\varepsilon)^n$, and thus vanishes as $n\to\infty$. Since this holds for all $\varepsilon>0$, and $\|S-S^\varepsilon\|_{cb}\le 2\,\varepsilon$, we have found a forgetful channel S^ε arbitrarily close to S, completing the proof. ▲

We will now show that the set of forgetful quantum channels is open. So assume that we are given a forgetful memory channel $S\colon\mathcal{B}\otimes\mathcal{M}\to\mathcal{M}\otimes\mathcal{A}$. We will show that S has a finite-size neighborhood in which all memory channels are likewise forgetful. Clearly, from the definition of forgetfulness we can find $N\in\mathbb{N}$ and a quantum channel $\tilde{S}_N\colon\mathcal{M}\to\mathcal{A}^{\otimes N}$ such that $\|\hat{S}_N - \mathbb{1}_{\mathcal{M}}\otimes\tilde{S}_N\|_{cb} < \frac{1}{2}$. Thus, for all memory channels T such that $\|T-S\|_{cb}\le\frac{1}{2N}$ we have

$$\|\hat{T}_N - \mathbb{1}_{\mathcal{M}}\otimes\tilde{S}_N\|_{cb} \le \|\hat{S}_N - \mathbb{1}_{\mathcal{M}}\otimes\tilde{S}_N\|_{cb} + N\,\|T-S\|_{cb} < 1\,, \tag{6.64}$$

and the forgetfulness of T immediately follows from Prop. 6.13. ■

It is instructive to observe that a forgetful channel is obtained from a possibly non-forgetful one in the denseness proof of Th. 6.20 by adding a tiny amount of white noise. In real-world experiments, such noise will always be present at some level. Therefore, quantum channels encountered in the laboratory will generally be forgetful.

However, while every non-forgetful quantum channel can be approximated by a forgetful memory channel to arbitrary degree of accuracy, their capacities may be very different. As an example for such a discontinuity effect, consider the channel with a global classical switch introduced in Sec. 6.2.3. Let us assume that Alice and Bob face a situation in which Eve controls the initial memory state and completely jams the communication. Then adding a little bit of noise, as in the proof of Th. 6.20, will deprive Eve of her control of the initial memory, and may lead to a channel with positive transmission rate. Thus, adding noise may actually be beneficial sometimes. Of course, it is just as easy to construct examples of memory channels which are rendered useless by adding a tiny amount of noise.

6.5 Forgetfulness and Cluster Properties

In this Section we will apply some results from quantum ergodic theory to prove just another interesting property of forgetful memory channels: they turn ergodic input states into ergodic output states. Ergodic states are those states that are extremal among all translational invariant states. As explained in Sec. 7.1.2 and Appendix C, they can in some sense be seen as a generalization of i.i.d. product states. In particular, the asymptotic equipartition properties that lie at the heart of all coding theorems for memoryless channels can be generalized to ergodic states [BKS$^+$04, BKS$^+$03]. Our coding schemes for forgetful channels presented in Sec. 6.6.2 do not rely on these generalizations, since they are based on a suitable approximation of forgetful channels by memoryless channels. However, an alternative approach that makes direct use of asymptotic equipartition for ergodic states (but so far only works for a limited class of quantum channels) has recently been suggested by Bjelaković and Boche [BB06].

Theorem 6.21. (Forgetfulness Preserves Ergodicity)
Let $T\colon \mathcal{B}_{\mathbb{Z}} \to \mathcal{A}_{\mathbb{Z}}$ be a forgetful causal channel in the sense of Def. 6.11, and suppose that $\omega \in \mathcal{A}_{\mathbb{Z}}^*$ is ergodic. Then $\omega \circ T \in \mathcal{B}_{\mathbb{Z}}^*$ is likewise ergodic.

The proof of Th. 6.21 is based on a useful characterization of ergodicity in terms of so-called *mixing* or *cluster properties*: a state $\omega \in \mathcal{A}_{\mathbb{Z}}^*$ is ergodic iff all observables that are located sufficiently far apart on the spin chain approximately factorize:

Proposition 6.22. (Cluster Properties)
A quantum state $\omega \in \mathcal{A}_{\mathbb{Z}}^*$ is ergodic iff

$$\inf_{z\in\mathbb{Z}} |\omega\big(X\,\sigma^z(Y)\big) - \omega(X)\,\omega\big(\sigma^z(Y)\big)| = 0 \tag{6.65}$$

for all $X, Y \in \mathcal{A}_{\mathbb{Z}}$, where σ denotes the shift operator (cf. App. C).

The proof of Prop. 6.22 relies on the Alaoglu-Birkhoff mean ergodic theorem and can be found as Ths. 4.3.17 and 4.3.22 in [BR87]. As explained there, it applies similarly to

states on C^*-algebras $\mathcal{A}$ which are ergodic with respect to a group of *-automorphisms. Note that asymptotic Abelianness is automatically satisfied for states on quasi-local algebras, since $\inf_{z\in\mathbb{Z}} \|[\sigma^z(X) \mid Y]\|_\infty = 0$ holds for all $X, Y \in \mathcal{A}_\mathbb{Z}$. We will now apply Prop. 6.22 in the

Proof of Th. 6.21: Let $X, Y \in \mathcal{B}_\mathbb{Z}$, and $\varepsilon > 0$. From the definition of the quasi-local algebra $\mathcal{B}_\mathbb{Z}$ as a norm-completion of local observables (cf. Appendix C), we know that there exist two finite subsets $\Lambda_X, \Lambda_Y \subset \mathbb{Z}$ and bounded operators $X_\varepsilon \in \mathcal{B}_{\Lambda_X}$ and $Y_\varepsilon \in \mathcal{B}_{\Lambda_Y}$ such that

$$\|X - X_\varepsilon\|_\infty \leq \varepsilon \qquad \text{and} \qquad \|Y - Y_\varepsilon\|_\infty \leq \varepsilon\,. \tag{6.66}$$

Since the causal channel T is assumed to be forgetful, it immediately follows from Def. 6.11 that we may find an integer $z_1 \in \mathbb{Z}$ such that

$$\|T\big(X_\varepsilon\, \sigma_\mathcal{B}^{z_1}(Y_\varepsilon)\big) - T\big(X_\varepsilon\big)\, T\big(\sigma_\mathcal{B}^{z_1}(Y_\varepsilon)\big)\|_\infty \leq \varepsilon\,. \tag{6.67}$$

Noting that $\|T\|_\infty = 1 = \|\sigma_\mathcal{B}^z\|_\infty$ and $\|X_\varepsilon\|_\infty \leq \|X\|_\infty + \varepsilon$, we can now apply the triangle inequality to conclude from Eqs. (6.66) and (6.67) that

$$\begin{aligned}
&\|T\big(X\, \sigma_\mathcal{B}^{z_1}(Y)\big) - T\big(X\big)\, T\big(\sigma_\mathcal{B}^{z_1}(Y)\big)\|_\infty \\
&\qquad \leq \|T\big(X\, \sigma_\mathcal{B}^{z_1}(Y)\big) - T\big(X_\varepsilon\, \sigma_\mathcal{B}^{z_1}(Y)\big)\|_\infty \\
&\qquad\qquad + \|T\big(X_\varepsilon\, \sigma_\mathcal{B}^{z_1}(Y)\big) - T\big(X_\varepsilon\, \sigma_\mathcal{B}^{z_1}(Y_\varepsilon)\big)\|_\infty \\
&\qquad\qquad + \|T\big(X_\varepsilon\, \sigma_\mathcal{B}^{z_1}(Y_\varepsilon)\big) - T\big(X_\varepsilon\big)\, T\big(\sigma_\mathcal{B}^{z_1}(Y_\varepsilon)\big)\|_\infty \\
&\qquad\qquad + \|T\big(X_\varepsilon\big)\, T\big(\sigma_\mathcal{B}^{z_1}(Y_\varepsilon)\big) - T\big(X_\varepsilon\big)\, T\big(\sigma_\mathcal{B}^{z_1}(Y)\big)\|_\infty \\
&\qquad\qquad + \|T\big(X_\varepsilon\big)\, T\big(\sigma_\mathcal{B}^{z_1}(Y)\big) - T\big(X\big)\, T\big(\sigma_\mathcal{B}^{z_1}(Y)\big)\|_\infty \\
&\qquad \leq \|Y\|_\infty \|X - X_\varepsilon\|_\infty + \|X_\varepsilon\|_\infty \|Y - Y_\varepsilon\|_\infty + \varepsilon \\
&\qquad\qquad + \|X_\varepsilon\|_\infty \|Y - Y_\varepsilon\|_\infty + \|X - X_\varepsilon\|_\infty \|Y\|_\infty \\
&\qquad \leq 2\,\|Y\|_\infty\, \varepsilon + 2\,\big(\|X\|_\infty + \varepsilon\big)\,\varepsilon + \varepsilon\,.
\end{aligned} \tag{6.68}$$

Since $\omega \in \mathcal{A}_\mathbb{Z}^*$ is ergodic, Prop. 6.22 implies that we may $z_2 \in \mathbb{Z}$ such that

$$\begin{aligned}
&|\omega\Big(T\big(X\big)\, T\big(\sigma_\mathcal{B}^{z_2}(Y)\big)\Big) - \omega\Big(T\big(X\big)\Big)\, \omega\Big(T\big(\sigma_\mathcal{B}^{z_2}(Y)\big)\Big)| \\
&\qquad = |\omega\Big(T\big(X\big)\, \sigma_\mathcal{A}^{z_2}\big(T(Y)\big)\Big) - \omega\Big(T\big(X\big)\Big)\, \omega\Big(\sigma_\mathcal{A}^{z_2}\big(T(Y)\big)\Big)| \leq \varepsilon\,,
\end{aligned} \tag{6.69}$$

where we have used that T is translational invariant, $T \circ \sigma_\mathcal{B}^{z_2} = \sigma_\mathcal{A}^{z_2} \circ T$. Then, for $z \in \mathbb{Z}$ with $|z| \geq \max\{|z_1|, |z_2|\}$, another application of the triangle inequality allows us to infer from Eqs. (6.68) and (6.69) that

$$\begin{aligned}
|\omega\Big(T\big(X\, \sigma_\mathcal{B}^{z}(Y)\big)\Big) \;&-\; \omega\Big(T\big(X\big)\Big)\, \omega\Big(T\big(\sigma_\mathcal{B}^{z}(Y)\big)\Big)| \\
&\leq\; |\omega\Big(T\big(X\, \sigma_\mathcal{B}^{z}(Y)\big)\Big) - \omega\Big(T\big(X\big)\, T\big(\sigma_\mathcal{B}^{z}(Y)\big)\Big)| \\
&\qquad + |\omega\Big(T\big(X\big)\, T\big(\sigma_\mathcal{B}^{z}(Y)\big)\Big) - \omega\Big(T\big(X\big)\Big)\, \omega\Big(T\big(\sigma_\mathcal{B}^{z}(Y)\big)\Big)| \\
&\leq\; 2\,\|Y\|_\infty\, \varepsilon + 2\,\big(\|X\|_\infty + \varepsilon\big)\,\varepsilon + 2\,\varepsilon\,.
\end{aligned} \tag{6.70}$$

Since $\varepsilon > 0$ is arbitrary, Prop. 6.22 implies that $\omega \circ T$ is ergodic, as suggested. ■

6.6 Entropic Bounds and Channel Coding

In Secs. 6.2.3 and 6.2.4 we have computed the channel capacity of some interesting model channels. Now we will be concerned with statements that apply more generally. In Sec. 6.6.1 we will give entropic upper bounds on the capacity for classical and quantum information transfer. In Sec. 6.6.2 achievability of these bounds will be demonstrated for the important class of forgetful quantum channels.

6.6.1 Entropic Bounds

It has already been pointed out by Bowen and Mancini [BM04] that the standard mutual information bound (or *Holevo bound*) [Hol73] on the classical channel capacity as well as the coherent information bound [BNS98, BST98, BKN00, Dev05] on the quantum capacity can be extended to quantum channels with memory. In fact, these bounds ultimately depend only on the mutual information between Alice's input register and Bob's output register, and are independent of the internal structure of the quantum channel that connects both parties. The proofs familiar from the memoryless setting can therefore be directly applied to memory channels. They provide entropic upper bounds on the classical and quantum capacity of a quantum memory channel in all the four different settings discussed in Def. 6.1 — in terms of the Holevo bound χ and the coherent information I_c, respectively (cf. App. A).

Proposition 6.23. (Upper Bounds on the Classical Channel Capacity) *Let $\mathcal{H}_A$, $\mathcal{H}_B$, and $\mathcal{H}_M$ be Hilbert spaces, and let S_{n*} be the n-fold concatenation of a quantum memory channel $S_*\colon \mathcal{B}_*(\mathcal{H}_M) \otimes \mathcal{B}_*(\mathcal{H}_A) \to \mathcal{B}_*(\mathcal{H}_B) \otimes \mathcal{B}_*(\mathcal{H}_M)$. The classical information capacities of S are bounded from above as follows:*

$$C_{AB}(S) \leq \overline{\lim_{n\to\infty}} \frac{1}{n} \max_{\{p_i,\varrho_i\}} \chi(S_{n*}, \{p_i, \varrho_i\}), \tag{6.71}$$

$$C_{AE}(S) \leq \overline{\lim_{n\to\infty}} \frac{1}{n} \max_{\{p_i,\varrho_i\}} \chi(\mathrm{tr}_{\mathcal{M}} \circ S_{n*}, \{p_i, \varrho_i\}), \tag{6.72}$$

$$C_{EB,\mu}(S) \leq \overline{\lim_{n\to\infty}} \frac{1}{n} \max_{\{p_i,\varrho_i\}} \chi(S_{n*}, \{p_i, \mu \otimes \varrho_i\}), \tag{6.73}$$

$$C_{EE,\mu}(S) \leq \overline{\lim_{n\to\infty}} \frac{1}{n} \max_{\{p_i,\varrho_i\}} \chi(\mathrm{tr}_{\mathcal{M}} \circ S_{n*}, \{p_i, \mu \otimes \varrho_i\}), \tag{6.74}$$

where $\mu \in \mathcal{B}_*(\mathcal{H}_M)$ is Eve's initial memory state. If $d_M := \dim \mathcal{H}_M < \infty$, the bounds in Eqs. (6.71) and (6.72), and in Eqs. (6.73) and (6.74) coincide pairwise. If the channel S is forgetful, the bounds in Eqs. (6.71) and (6.73), and in Eqs. (6.72) and (6.74) coincide pairwise.

Proposition 6.24. (Upper Bounds on the Quantum Channel Capacity) *The quantum information capacities of the memory channel S are bounded from above as follows:*

$$Q_{AB}(S) \leq \overline{\lim_{n\to\infty}} \frac{1}{n} \max_{\varrho} I_c(S_{n*}, \varrho)\,, \tag{6.75}$$

$$Q_{AE}(S) \leq \overline{\lim_{n\to\infty}} \frac{1}{n} \max_{\varrho} I_c(\mathrm{tr}_{\mathcal{M}} \circ S_{n*}, \varrho)\,, \tag{6.76}$$

$$Q_{EB,\mu}(S) \leq \overline{\lim_{n\to\infty}} \frac{1}{n} \max_{\varrho} I_c(S_{n*}, \mu \otimes \varrho)\,, \tag{6.77}$$

$$Q_{EE,\mu}(S) \leq \overline{\lim_{n\to\infty}} \frac{1}{n} \max_{\varrho} I_c(\mathrm{tr}_{\mathcal{M}} \circ S_{n*}, \mu \otimes \varrho)\,, \tag{6.78}$$

where $\mu \in \mathcal{B}_*(\mathcal{H}_M)$ is Eve's initial memory state. If $d_M < \infty$, the bounds in Eqs. (6.75) and (6.76), and in Eqs. (6.77) and (6.78) coincide pairwise. If the channel S is forgetful, the bounds in Eqs. (6.75) and (6.77), and in Eqs. (6.76) and Eq. (6.78) coincide pairwise.

Remark 6.25. Note that the bounds in Props. 6.23 and 6.24 still hold when we only require that coding is possible along some (possibly very sparse) block sequence $(n_\nu)_{\nu\in\mathbb{N}}$. In Def. 6.1 we have been more ambitious, since we have required that coding works for arbitrary block sizes. When this stronger version of capacity is chosen, the $\overline{\lim}$ can be replaced by $\underline{\lim}$ in Eqs. (6.71) through (6.78). While the "optimistic" and the "pessimistic" channel capacity coincide for memoryless channels, this is not clear for channels with memory (cf. Remark 6.5). For forgetful channels, equivalence does hold, as will be seen in Sec. 6.6.2.

Proof of Props. 6.23 and 6.24: As indicated above, the proof transfers directly from the memoryless setting. We thus refer to Holevo's original work [Hol73] for the classical bound, and to the works of Barnum *et al.* [BNS98, BST98, BKN00] and Devetak [Dev05] for the quantum case.

Here we only show that the bounds coincide pairwise under the additional assumption of having a memory of finite size or a forgetful channel. We will begin with the finite memory case: Since the Holevo quantity χ decreases under quantum operations (cf. App. A), we find

$$\chi(\mathrm{tr}_{\mathcal{M}} \circ S_{n*}, \{p_i, \varrho_i\}) \leq \chi(S_{n*}, \{p_i, \varrho_i\}) \leq \chi(\mathrm{tr}_{\mathcal{M}} \circ S_{n*}, \{p_i, \varrho_i\}) + 2\,\mathrm{ld}\, d_M\,, \tag{6.79}$$

where in the last step the subadditivity of the von Neumann entropy has been applied [NC00]. From Eq. (6.79) it immediately follows that the bounds on C_{AB} and C_{AE} coincide whenever $d_M < \infty$. The proof for the bounds on $C_{EB,\mu}$ and $C_{EE,\mu}$ is completely analogous.

For the bounds on the quantum capacities, we use the *data processing inequality* (cf. App. A) and again apply the subadditivity of the von Neumann entropy. ▲

Now we assume that S is forgetful. Prop. 6.17 entails that for any $\varepsilon > 0$ we may find a positive integer $m \in \mathbb{N}$ such that

$$\|\mathrm{tr}_{\mathcal{B}^{\otimes m}} S_{m*}\big(\varrho_1 - \varrho_2\big)\|_1 \leq \varepsilon \tag{6.80}$$

for all density operators $\varrho_1, \varrho_2 \in \mathcal{B}_*(\mathcal{H}_M) \otimes \mathcal{B}_*(\mathcal{H}_A)^{\otimes m}$ satisfying $\mathrm{tr}_{\mathcal{M}}\varrho_1 = \mathrm{tr}_{\mathcal{M}}\varrho_2$. Applying Fannes' Inequality (cf. App. A) and the subadditivity of the von Neumann entropy, we can thus conclude that for arbitrary $\mu \in \mathcal{B}_*(\mathcal{H}_M)$ and $n \in \mathbb{N}$ we have

$$\begin{aligned}\chi(S_{n*}, \{p_i, \varrho_i\}) &\leq \chi(\mathrm{tr}_{\mathcal{B}^{\otimes m}} S_{n*}, \{p_i, \varrho_i\}) + 2\,m \operatorname{ld} d_B \\ &\leq \chi(\mathrm{tr}_{\mathcal{B}^{\otimes m}} S_{n*}, \{p_i, \mu \otimes \mathrm{tr}_{\mathcal{M}}(\varrho_i)\}) + 2\,m \operatorname{ld} d_B \\ &\qquad + \frac{1}{e\,\ln 2} + 2\,\|\mathrm{tr}_{\mathcal{B}^{\otimes m}} S_{n*}\big(\varrho_i - \mu \otimes \mathrm{tr}_{\mathcal{M}}(\varrho_i)\big)\|_1 \operatorname{ld} d_B^n \\ &\leq \max_{\{q_j, \sigma_j\}} \chi(S_{n*}, \{q_j, \mu \otimes \sigma_j\}) + 2\,m \operatorname{ld} d_B + \frac{1}{e\,\ln 2} + 2\,n\,\varepsilon \operatorname{ld} d_B\,.\end{aligned} \tag{6.81}$$

Maximizing over the ensemble $\{p_i, \varrho_i\}$, dividing by n and letting $n \to \infty$, we may conclude from Eq. (6.81) that

$$\overline{\lim_{n\to\infty}} \frac{1}{n} \max_{\{p_i, \varrho_i\}} \chi(S_{n*}, \{p_i, \varrho_i\}) \leq \overline{\lim_{n\to\infty}} \frac{1}{n} \max_{\{p_i, \varrho_i\}} \chi(S_{n*}, \{p_i, \mu \otimes \varrho_i\}) + 2\,\varepsilon \operatorname{ld} d_B\,, \tag{6.82}$$

implying that for every $\mu \in \mathcal{B}_*(\mathcal{H}_M)$ the bound on the classical capacity $C_{EB,\mu}$ is no smaller than the bound on the capacity C_{AB}. The converse estimate is immediate, since Alice can obviously choose quantum ensembles of the form $\{p_i, \mu \otimes \varrho_i\}$ if she has access to the input memory. The proof for the bounds on $C_{EE,\mu}$ and C_{AE} is completely analogous, as is the proof for the quantum case. ■

6.6.2 Coding Theorems for Forgetful Channels

In this Section we will demonstrate that for forgetful channels the entropic bounds on the classical and quantum channel capacities presented in Prop. 6.23 and Prop. 6.24 are in fact achievable rates, and the limits exist.

The idea for the proof is a reduction of the problem to the memoryless setting via a relatively simple double-blocking procedure. To illustrate the strategy, let's start with the easy case in which there is a finite integer $m \in \mathbb{N}$ such that

$$\hat{S}_m = (P \otimes \mathrm{id}_{\mathcal{A}}^{\otimes m}) \circ \hat{S}_m\,, \tag{6.83}$$

where $P\colon \mathcal{M} \to \mathbb{C}\,\mathbb{1}_{\mathcal{M}}$ is again the completely depolarizing channel. We call channels with this property *strictly forgetful*, and the smallest integer m such that Eq. (6.83) is satisfied will be called the *memory depth* of the channel S. For the processing of long messages, we group the channels into blocks of length $m + l$ and ignore the outputs of the first m channels of each block, while the actual coding is done for the remaining l

channels. Eventually we will let $l \to \infty$. When we restrict the inputs to product states of block length $m + l$, due to strict forgetfulness the output state factorizes, and the whole setup corresponds to a memoryless channel on the larger input space $\mathcal{H}_A^{\otimes l+m}$. For the transmission of classical information, we can then apply the standard HSW random coding techniques [Hol98, SW97]. Invoking subadditivity of von Neumann entropy as in Sec. 6.6.1, the rates r which can be achieved with this coding scheme are seen to be bounded as follows:

$$\frac{1}{l+m} \max_{\{p_i,\varrho_i\}} \chi(S_{l*}, \{p_i, \varrho_i\}) - \frac{2\,m}{m+l} \operatorname{ld} d_B \leq r \leq \frac{1}{l} \max_{\{p_i,\varrho_i\}} \chi(S_{l*}, \{p_i, \varrho_i\})\,. \tag{6.84}$$

The claim now follows by letting $l \to \infty$. For quantum channel capacities, Devetak's coding theorem [Dev05], as sketched in Sec. 5.5.4, can be shown to yield an analogous bound, in which the Holevo quantity is replaced by the coherent information.

It turns out that we can apply the same double-blocking strategy even if the memory channel S is merely assumed to be forgetful (and no longer strictly forgetful). However, in this case the output states do not completely factorize, and the error we pick up by replacing the memory channel with a memoryless channel on larger blocks grows with the number of blocks. Luckily, all memory effects can be assumed to vanish exponentially fast by Cor. 6.16. We will show below that the double-blocking scheme then allows to transfer the coding theorems for the (private) classical and quantum channel capacities from the memoryless realm to forgetful quantum channels.

Theorem 6.26. (Coding for Forgetful Quantum Channels)
Let $\mathcal{H}_A$, $\mathcal{H}_B$, and $\mathcal{H}_M$ be finite-dimensional Hilbert spaces, and let us assume that $S_*\colon \mathcal{B}_*(\mathcal{H}_M) \otimes \mathcal{B}_*(\mathcal{H}_A) \to \mathcal{B}_*(\mathcal{H}_B) \otimes \mathcal{B}_*(\mathcal{H}_M)$ is a forgetful quantum channel. By S_{n*} we denote its n-fold concatenation, and by S_E its complementary channel in the sense of Sec. 3.3. With the convention introduced in Remark 6.3, we then have

$$C_*(S) = \lim_{n\to\infty} \frac{1}{n} \max_{\{p_i,\varrho_i\}} \chi(S_{n*}, \{p_i, \varrho_i\})\,, \tag{6.85}$$

$$C_*^p(S) = \lim_{n\to\infty} \frac{1}{n} \max_{\{p_i,\varrho_i\}} \Big\{ \chi(S_{n*}, \{p_i, \varrho_i\}) - \chi(S_{En*}, \{p_i, \varrho_i\}) \Big\}\,, \tag{6.86}$$

$$Q_*(S) = \lim_{n\to\infty} \frac{1}{n} \max_{\varrho} I_c(S_{n*}, \varrho)\,. \tag{6.87}$$

Proof: The proof of the upper bound on the private classical capacity $C_*^p(S)$, i. e.,

$$C_{AB}^p(S) \leq \overline{\lim_{n\to\infty}} \frac{1}{n} \max_{\{p_i,\varrho_i\}} \Big\{ \chi(S_{n*}, \{p_i, \varrho_i\}) - \chi(S_{En*}, \{p_i, \varrho_i\}) \Big\}\,, \tag{6.88}$$

is completely analogous to the one for the memoryless case [Dev05]. For $C_{AB}(S)$ and $Q_{AB}(S)$, corresponding results have been presented in Props. 6.23 and 6.24. To complete the proof it thus remains to show that

$$C_{EE,\mu}(S) \geq \lim_{n\to\infty} \frac{1}{n} \max_{\{p_i,\varrho_i\}} \chi(S_{n*}, \{p_i, \varrho_i\}) \tag{6.89}$$

for all $\mu \in \mathcal{B}_*(\mathcal{H}_M)$, and that the limit on the RHS of Eq. (6.89) exists, and correspondingly for $C^p_{EE,\mu}(S)$ and $Q_{EE,\mu}(S)$.

The definition of forgetfulness combined with Cor. 6.16 implies that we may find a sequence $(\tilde{S}_m)_{m\in\mathbb{N}}$ of quantum channels such that

$$\|\hat{S}_m - \mathbb{1}_{\mathcal{M}} \otimes \tilde{S}_m\|_{cb} \leq c^{-m} \tag{6.90}$$

for some constant $c > 1$. As described above for the case of strictly forgetful channels, our strategy is then to group the memory channels into blocks of length $m+l$, to ignore the outputs on the first m channels of each block, and to replace the resulting channel $T_{m+l} := (\hat{S}_m \otimes \mathrm{id}_{\mathcal{A}^{\otimes l}}) \circ S_l$ with the memoryless channel

$$\tilde{T}_{m+l} := (\mathbb{1}_{\mathcal{M}} \otimes \tilde{S}_m \otimes \mathrm{id}_{\mathcal{A}^{\otimes l}}) \circ S_l \,. \tag{6.91}$$

For Alice, this coding procedure means that she will have to feed the first m inputs of each block of length $m+l$ with some standard state $\omega \in \mathcal{B}_*(\mathcal{H}_A)^{\otimes m}$, while she will use the remaining l inputs of each block for the actual coding. Bob will ignore the first m output signals of each block, and will run his decoding algorithm on the remaining l signals.

Let us focus on the classical information capacity first, and assume that we have a coding scheme for the memoryless channel $\tilde{T}_{m+l}$ that achieves the rate r. According to Def. 5.4, this means that for every $\varepsilon > 0$ there is an integer $N_\varepsilon \in \mathbb{N}$ such that for every $n \geq N_\varepsilon$ we may find a code book with $\nu := \lfloor 2^{nlr} \rfloor$ codewords $\{\varrho_j\}_{j=1}^{\nu} \subset \mathcal{B}_*(\mathcal{H}_A)^{\otimes ln}$ and a corresponding observable $\{M_j\}_{j=1}^{\nu} \subset \mathcal{B}(\mathcal{H}_B)^{\otimes ln}$ such that

$$\mathrm{tr}\, \tilde{T}^{\otimes n}_{m+l*}(\varrho_j)\, M_j \;\geq\; 1-\varepsilon \qquad \forall\, n \geq N_\varepsilon \,, \tag{6.92}$$

uniformly in $j = 1, ..., \nu$. By the results of Holevo [Hol98] and Schumacher and Westmoreland [SW97], such coding schemes exist for all rates $r < \frac{l}{m+l} C_1(\tilde{T}_l)$, where $C_1(\tilde{T}_l)$ denotes the product state capacity of the memoryless channel $\tilde{T}_l$.

As explained in Sec. 5.5.4, for the private classical information capacity the setting is basically the same, but the codewords $\{\varrho_{jk}\}_{j=1,k=1}^{\nu_B,\nu_E}$ carry a second index to allow for randomization, and there exists an operator $\Theta \in \mathcal{B}(\mathcal{H}_E)^{\otimes nl}$ such that

$$\Big\| \frac{1}{\nu_E} \sum_{k=1}^{\nu_E} \tilde{T}^{\otimes n}_{El*}(\varrho_{jk}) \;-\; \Theta \Big\|_1 \leq \varepsilon \qquad \forall\, j = 1, ..., \nu_B \,. \tag{6.93}$$

In this case the size of the code is given by $\nu_B = \lfloor 2^{nlr} \rfloor$, and all rates $r < \frac{l}{l+m} C_1^p(\tilde{T}_l)$ are achievable.

The same product coding scheme will now be applied to the concatenated memory channel T_{m+l}. Our objectives are to show that

(*a*) this coding scheme satisfies the decoding condition Eq. (6.92),

(b) in the case of private information transfer, the privacy condition Eq. (6.93) holds, and

(c) the attainable rates can be brought arbitrarily close to the entropic upper bounds.

This will immediately imply the coding theorem for classical and private classical information transfer. The quantum channel coding theorem will then follow from the coherentification of the private classical protocol, as explained in detail in Devetak's original work [Dev05] and sketched in Sec. 5.5.4. In fact, the coherentification protocol itself applies generally and does not depend on the internal structure of the quantum channel that connects the sender to the receiver and the environment.

Let us start with the decoding condition (a). Assume that in n blocks of length $m+l$ each, the replacement $T_{m+l} \mapsto \tilde{T}_{m+l}$ is made. Since $\|T_{m+l} - \tilde{T}_{m+l}\|_{cb} \leq c^{-m}$ for each of these blocks by Eq. (6.90), the concatenated channels satisfy

$$\|T_{n(m+l)} - \tilde{T}_{m+l}^{\otimes n}\|_{cb} \;\leq\; n\, c^{-m}. \tag{6.94}$$

Making use of the norm duality Eq. (6.50), we can conclude from Eq. (6.94) that

$$\|T_{n(m+l)*}(\varrho) - \tilde{T}_{m+l*}^{\otimes n}(\varrho)\|_1 \leq n\, c^{-m}. \tag{6.95}$$

for all input states ϱ. Noting that for any two quantum states $\varrho, \sigma \in \mathcal{B}_*(\mathcal{H})$ and any observable $\{M_j\}_{j=1}^{\nu} \subset \mathcal{B}(\mathcal{H})$ the inequality

$$\|\varrho - \sigma\|_1 \geq \sum_{j=1}^{\nu} |\mathrm{tr}\, M_j\, (\varrho - \sigma)| \tag{6.96}$$

holds true (cf. [NC00], Th. 9.1), we may infer from Eq. (6.95) that for all codewords $\{\varrho_j\}_{j=1}^{\nu} \subset \mathcal{B}_*(\mathcal{H}_A)^{\otimes ln}$ we have

$$\begin{aligned} \mathrm{tr}\, T_{n(m+l)*}(\varrho_j)\, M_j &\geq \mathrm{tr}\, \tilde{T}_{m+l*}^{\otimes n}(\varrho_j)\, M_j - \|T_{n(m+l)*}(\varrho_j) - \tilde{T}_{m+l*}^{\otimes n}(\varrho_j)\|_1 \\ &\geq \mathrm{tr}\, \tilde{T}_{m+l*}^{\otimes n}(\varrho_j)\, M_j - n\, c^{-m}. \end{aligned} \tag{6.97}$$

For $\varepsilon > 0$, we now choose $n := l$, $m := \varepsilon\, l$, and l sufficiently large such that Eq. (6.92) is satisfied. We may then conclude from Eq. (6.97) that

$$\mathrm{tr}\, T_{l^2(1+\varepsilon)*}(\varrho_j)\, M_j \;>\; 1 - 2\,\varepsilon \tag{6.98}$$

uniformly in j for sufficiently large l, implying that the product channel random coding scheme leads to asymptotically vanishing errors for all transfer rates $r < \frac{1}{1+\varepsilon}\, C_1(\tilde{T}_l)$ and $r < \frac{1}{1+\varepsilon}\, C_1^p(\tilde{T}_l)$, respectively.

We will now show that (b) also holds, with the same substitution $\varepsilon \mapsto 2\varepsilon$. To this end, we note that Devetak's randomization scheme can be slightly modified to include the output memory state of each block. This trick guarantees that in an l-fold concatenation

of blocks of length $m+l$ each, even the intermediate blocks, for which no coding is done and the respective outputs are ignored, are (almost) uncorrelated with Alice's signal states.

Making again use of the error estimate for concatenated channels and the norm duality Eq. (6.50), we may then conclude from Eq. (6.93) that

$$\begin{aligned}\Big\|\frac{1}{\nu_E}\sum_{k=1}^{\nu_E} T_{E\,l(m+l)*}\left(\varrho_{jk}\right) - \Theta\Big\|_1 &\leq \Big\|\frac{1}{\nu_E}\sum_{k=1}^{\nu_E}\Big[T_{E\,l(m+l)*}\left(\varrho_{jk}\right) - \tilde{T}_{E\,m+l*}^{\otimes l}\left(\varrho_{jk}\right)\Big]\Big\|_1 \\ &\quad + \Big\|\frac{1}{\nu_E}\sum_{k=1}^{\nu_E}\tilde{T}_{E\,m+l*}^{\otimes l}\left(\varrho_{jk}\right) - \Theta\Big\|_1 \\ &\leq l\,c^{-m} + \varepsilon \\ &= l\,c^{-\varepsilon l} + \varepsilon \\ &\leq 2\,\varepsilon \end{aligned} \tag{6.99}$$

for sufficiently large l, as advertised. Note that without the additional randomization over the output memory, the average mutual information $\frac{1}{l^2}H(A:E)$ between the signal states and Eve's output states will still be small. This is due to the fact that in the above coding scheme the intermediate blocks only constitute a fraction ε of the total length. However, this is in general not sufficient to conclude that a norm estimate such as Eq. (6.99) holds.

In order to conclude the proof, it only remains to show that $C_1(\tilde{T}_l)$ can be bounded from below in terms of $\max_{\{p_i,\varrho_i\}}\chi(S_{l*},\{p_i,\varrho_i\})$ for large l, and similarly for the private classical and quantum capacities. Applying the subadditivity of the von Neumann entropy and Fannes' inequality (cf. App. A), we see that

$$\begin{aligned}\chi(S_{l*},\{p_i,\varrho_i\}) &\leq \chi(T_{l+\varepsilon l*},\{p_i,\varrho_i\}) + 2\,\varepsilon\,l\,\mathrm{ld}\,d_B \\ &\leq \chi(\tilde{T}_{l+\varepsilon l*},\{p_i,\varrho_i\}) + 2\,\varepsilon\,l\,\mathrm{ld}\,d_B + \frac{2}{e\,\ln 2} + 2\,l\,(1+\varepsilon)\,\varepsilon\,\mathrm{ld}\,d_B \\ &\leq l\,(1+\varepsilon)\,C_1(\tilde{T}_l) + 2\,\varepsilon\,l\,\mathrm{ld}\,d_B + \frac{2}{e\,\ln 2} + 2\,l\,(1+\varepsilon)\,\varepsilon\,\mathrm{ld}\,d_B\,. \end{aligned} \tag{6.100}$$

Since $C_1(\tilde{T}_l)$ has been shown to be an achievable rate for large enough l, we may conclude from Eq. (6.100) that

$$C_{EE,\mu}(S) \geq \frac{1}{1+\varepsilon}\Big[\varlimsup_{l\to\infty}\frac{1}{l}\max_{\{p_i,\varrho_i\}}\chi(S_{l*},\{p_i,\varrho_i\}) - 4\,\varepsilon\,\mathrm{ld}\,d_B - 2\,\varepsilon^2\,\mathrm{ld}\,d_B\Big]\,. \tag{6.101}$$

Since $\varepsilon > 0$ is arbitrary, Eq. (6.101) together with the upper bound in Prop. 6.23 entails that

$$C_{EE,\mu}(S) = \varlimsup_{n\to\infty}\frac{1}{n}\max_{\{p_i,\varrho_i\}}\chi(S_{n*},\{p_i,\varrho_i\})\,. \tag{6.102}$$

The coding scheme described above uses blocks of length $n_l := l^2(1+\varepsilon)$. This is a subexponential sequence in the sense of Remark 6.5, and we may thus apply the

one-sequence theorem from Sec. 5.2.1 to conclude that the limit in Eq. (6.102) exists, implying that Eq. (6.85) holds. The rate estimate for the private classical and quantum capacities is completely analogous. ■

6.7 Summary and Outlook

In this Chapter we have presented a general model for quantum channels with memory, and shown that under natural causality constraints every quantum process can be thought of as a concatenated memory channel (plus some memory initializer).

For these memory channels, channel capacities have been introduced along the lines familiar from the memoryless context, and it has been demonstrated that different operational setups may lead to different values of the channel capacity.

While we have concentrated on the classical and quantum channel capacities proper, it is evident that the theory may be extended to memory channels assisted by additional resources, such as entanglement and classical side communication (cf. Sec. 5.4.2). As explained in Sec. 6.6.1, entropic bounds typically depend only on the amount of information shared by sender and receiver, and not on the internal structure of the quantum channel connecting these two. Coding theorems for memoryless channels can easily be extended to forgetful memory channels, as demonstrated in Sec. 6.6.2. They typically lead to regularized expressions for the channel capacity, which still require the solution of optimization problems in Hilbert spaces of exponentially growing dimensionality. In general, computing capacities of quantum memory channels is thus at least as challenging as for memoryless channels, with less hope for improvements.

A complete investigation of the resulting capacity landscape is still pending. In particular, we do not yet know under which general conditions some (or all) of the channel capacities introduced in Def. 6.1 coincide. It may seem reasonable to conjecture that, as long as the memory system is finite-dimensional, it is irrelevant for capacity purposes whether Bob or Eve control the final memory output. While this is almost immediate for the entropic upper bounds on the channel capacities (cf. Prop. 6.23 and Prop. 6.24), so far we have not been able to verify this conjecture for the capacities themselves.

We have demonstrated in Sec. 6.4.4 that generic memory channels are forgetful, and in Sec. 6.6.2 we have presented coding theorems for this very important class of channels. It is thus tempting to conclude that it should be possible to always restrict one's attention to forgetful channels. However, the capacity of a memoryless channel is sometimes discontinuous in its parameters. So while it is always possible to approximate a given non-forgetful channel by a forgetful channel to arbitrary degree of accuracy, their capacities may be very different, as the example given in Sec. 6.4.4 demonstrates. This calls for a more detailed analysis of non-forgetful quantum channels and their capacities.

While we have presented several equivalent criteria for a memory channel to be forgetful (cf. Sec. 6.4.1), we do not yet have a structure theorem to characterize all the non-forgetful quantum channels, nor do we have a simple test to decide whether a given memory channel is forgetful.

Apart from some relatively simple model channels, little is known so far about the channel capacity of general non-forgetful memory channels. The derivation of coding theorems in this case is likely to require *universal* coding schemes, with encoders and decoders independent of Eve's choice of the initial memory state. For the memory channel with a global classical switch (cf. Sec. 6.2.3), universal coding schemes do exist — at least for the classical capacity [DD06]. However, this is a rather special example of a memory channel, and the general case remains very much open.

Chapter 7

Quantum Information Spectrum

This Chapter will serve as a brief introduction to quantum information spectrum methods as an approach to coding theorems for completely general quantum sources and channels. In contrast to the results on quantum memory channels described in Ch. 6, the information spectrum methods do not rely on asymptotic equipartition properties. As a consequence, structural assumptions such as stationarity, ergodicity, or even causality can all be dropped.

After a rather extensive introduction to the basic ideas in Sec. 7.1, we will define quantum spectral divergence rates as a generalization of the quantum relative entropy in Sec. 7.2 and explore some of their properties. In Sec. 7.3 we will then prove a source coding theorem for general quantum sources, and apply these results to ergodic as well as mixed sources. Sec. 7.4 summarizes the results and comments on possible future developments.

The work described in this Chapter arose from discussions with Garry Bowen and Nilanjana Datta.

7.1 Introduction and Motivation

As explained in Sec. 1.3, information theory is a discipline that tries to relate operational concepts such as the compression rate of a quantum source or the transmission rate of a quantum channel to information-theoretic quantities such as entropy or mutual information. Coding theorems then take on the form

$$OP = INF\,, \tag{7.1}$$

where OP is an operational quantity involving concepts such as encoding and decoding operations, key length, or asymptotic transmission rates, and INF is some information-theoretic quantity, typically the solution of some entropic extremal value problem.

7.1.1 Schumacher Compression Revisited

The fact that all the standard coding problems in classical as well as quantum information theory have a solution in terms of some entropic expression is a direct reflection of the law of large numbers, which entails asymptotic equipartition properties for probabilities and eigenvalues. The general concept and philosophy is nicely demonstrated by Schumacher's compression theorem for an i.i.d. quantum information source [Sch95, JS94]. The objective of quantum data compression is to determine the minimal physical requirements needed to store quantum information in such a way that it can later be faithfully reproduced. Rigorous definitions will be provided in Sec. 7.3; at this point we only sketch the basic ideas necessary to understand how the von Neumann entropy arises in the solution to this problem.

If the quantum information to be compressed is modelled as a quantum state $\varrho \in \mathcal{B}_*(\mathcal{H})$ on some d-dimensional Hilbert space $\mathcal{H}$, a compression scheme with rate r consists of a sequence of compression operations $(C_n)_{n\in\mathbb{N}}$ and decompression operations $(D_n)_{n\in\mathbb{N}}$ such that C_n maps the quantum state $\varrho^{\otimes n}$ to a 2^{nr}-dimensional subspace, from which $\varrho^{\otimes n}$ can be faithfully recovered by means of the decoding operation D_n. Our interest is to minimize the required resources, i.e., to find the smallest such compression rate r such that the probability of a failure in the reconstruction procedure vanishes asymptotically as the message length increases. As explained in Sec. 2.8, Schumacher [Sch95, JS94] was able to show that the smallest such compression rate is just the von Neumann entropy, $H(\varrho) = -\mathrm{tr}\,\varrho \,\mathrm{ld}\,\varrho$, and this is why we regard entropy as a suitable measure of quantum information.

In the following we will briefly sketch the central idea of Schumacher's proof to explain how entropic expressions arise from the equipartition of eigenvalues. A full proof of a corresponding theorem for completely general quantum sources will then be presented in Sec. 7.3.1.

Let $\varrho = \sum_{i=1}^{d} \lambda_i \,|\psi_i\rangle\langle\psi_i|$ be the spectral decomposition of the quantum state $\varrho \in \mathcal{B}_*(\mathcal{H})$. Then its n-fold tensor product can be given the representation

$$\varrho^{\otimes n} = \sum_{i_1,\ldots,i_n=1}^{d} \lambda_{i_1}\cdots\lambda_{i_n}\, |\hat{\psi}_{i_1,\ldots,i_n}\rangle\langle\hat{\psi}_{i_1,\ldots,i_n}|\,, \tag{7.2}$$

where we have set $|\hat{\psi}_{i_1,\ldots,i_n}\rangle := |\psi_{i_1}\rangle \otimes |\psi_{i_2}\rangle \cdots \otimes |\psi_{i_n}\rangle$, the product of the respective eigenvectors. For some (small) $\delta > 0$, we now define the *δ-typical* eigenvalues of $\varrho^{\otimes n}$ as those that are concentrated around $2^{-nH(\varrho)}$,

$$T_\delta^n := \left\{(i_1,\ldots,i_n) \;\mid\; \lambda_{i_1}\cdots\lambda_{i_n} \in [2^{-n(H(\varrho)+\delta)}, 2^{-n(H(\varrho)-\delta)}]\right\}. \tag{7.3}$$

To understand why these eigenvalues are called *typical*, recall that for any sequence $(X_i)_{i\in\mathbb{N}}$ of independent and identically distributed (i.i.d.) random variables with common finite expectation value $\mathbb{E}(X_1)$ and finite second moments, the weak law of large

numbers states that

$$\lim_{n\to\infty} \mathbb{P}\Big(\big|\frac{1}{n}\sum_{i=1}^{n} X_i - \mathbb{E}(X_1)\big| > \varepsilon\Big) = 0 \tag{7.4}$$

holds for every $\varepsilon > 0$. Applying this theorem to the random variable $\mathrm{ld}\,\lambda_i$, we immediately have

$$\frac{1}{n}\mathrm{ld}\,\lambda_{i_1}\cdots\lambda_{i_n} = \frac{1}{n}\sum_{j=1}^{n}\mathrm{ld}\,\lambda_{i_j} \xrightarrow[n\to\infty]{\mathbf{P}} \mathbb{E}(\mathrm{ld}\,\lambda_{i_1}) = \sum_{i_1=1}^{d}\lambda_{i_1}\,\mathrm{ld}\,\lambda_{i_1} = -H(\varrho)\,, \tag{7.5}$$

implying that asymptotically all eigenvalues of $\varrho^{\otimes n}$ are of order $\sim 2^{-nH(\varrho)}$, and hence fall into the typical subspace T_δ^n. The corresponding projector

$$P_\delta^n := \sum_{(i_1,\ldots,i_n)\in T_\delta^n} |\hat{\psi}_{i_1,\ldots,i_n}\rangle\langle\hat{\psi}_{i_1,\ldots,i_n}| \tag{7.6}$$

on the δ-typical subspace then almost fully supports the tensor product state $\varrho^{\otimes n}$ for large enough n,

$$\lim_{n\to\infty} \mathrm{tr}\; P_\delta^n\,\varrho^{\otimes n} = 1 \tag{7.7}$$

for any $\delta > 0$. We may therefore compress the quantum source $\varrho^{\otimes n}$ simply by projecting on the typical subspace, $C_n(\sigma) := P_\delta^n \sigma P_\delta^n$. The decoding operation can then be chosen as the identity channel $D_n := \mathrm{id}_n$ on the typical subspace, and Eq. (7.7) implies that the quantum state can be faithfully recovered from that subspace. The compression rate r for this scheme is given by the size of the typical subspace, $r = \frac{1}{n}\mathrm{ld}\,|T_\delta^n| = \frac{1}{n}\mathrm{ld}\,\mathrm{tr}\; P_\delta^n$. From the estimate

$$1 \geq \sum_{(i_1,\ldots,i_n)\in T_\delta^n} \lambda_{i_1}\cdots\lambda_{i_n} \geq |T_\delta^n|\, 2^{-n(H(\varrho)+\delta)} \tag{7.8}$$

we immediately find the upper bound

$$r = \frac{1}{n}\mathrm{ld}\,|T_\delta^n| \leq H(\varrho) + \delta\,, \tag{7.9}$$

which holds for arbitrary $\delta > 0$. Hence, we can compress the quantum source $\varrho^{\otimes n}$ down to at least the entropy $H(\varrho)$ by projecting on the typical subspace. Schumacher has also shown that no better scheme exists, but we will leave this point to Sec. 7.3.

7.1.2 Beyond Ergodicity

The sketch of the compression theorem for an i.i.d. quantum information source in Sec. 7.1.1 shows how the law of large numbers results in equipartition of the eigenvalues and an entropic expression for the optimal compression rate. All known coding theorems in classical and quantum information theory are proved along similar lines. The coding theorems for forgetful quantum memory channels presented in Sec. 6.6.2

also fall into this category, since for our proof we have approximated forgetful channels with memoryless channels on larger blocks.

The memoryless (i.i.d.) setting can sometimes be generalized to ergodicity (cf. Sec. 6.5 and Appendix C), where asymptotic equipartition properties continue to hold [BKS$^+$04, BKS$^+$03]. But how do we proceed if little or no structural assumptions on the nature of the source or channel are made, and both ergodicity and asymptotic equipartition will in general fail?

Information spectrum methods replace the idea of typical events with high probability events and provide a technique to deal with completely general sources and channels. As an illustrative example, let us again focus on quantum source compression — a preview of the problem that will be treated in detail in Sec. 7.3. Assume that we are given a sequence of Hilbert spaces $(\mathcal{H}_n)_{n\in\mathbb{N}}$ and a quantum source

$$\hat{\varrho} := (\varrho_1, \varrho_2, \cdots, \varrho_n, \cdots) \equiv (\varrho_n)_{n\in\mathbb{N}}\,, \tag{7.10}$$

where each $\varrho_n \in \mathcal{B}_*(\mathcal{H}_n)$ is a density operator. Note that at this point we do not make any structural assumptions on the source $\hat{\varrho}$. We do not even require a consistency condition such as $\mathrm{tr}_j \varrho_j = \varrho_{j-1}$, even though this may seem a completely natural demand in order for $\hat{\varrho}$ to have an interpretation as a physical information source. We now define the *spectral sup-entropy rate* of $\hat{\varrho}$ as

$$\overline{H}(\hat{\varrho}) := \inf\left\{\gamma \;\middle|\; \overline{\lim_{n\to\infty}}\ \mathrm{tr}\,\{\varrho_n \geq 2^{-n\gamma}\}\varrho_n = 1\right\}, \tag{7.11}$$

where for any self-adjoint operator A with spectral decomposition $A = \sum_i \lambda_i\, |i\rangle\langle i|$ and any real number $\beta \in \mathbb{R}$ we set

$$\{A \geq \beta\} := \sum_{\lambda_i \geq \beta} |i\rangle\langle i|\,. \tag{7.12}$$

We will show in Sec. 7.3.1 that $\overline{H}(\hat{\varrho})$ is the optimal compression rate of the quantum source $\hat{\varrho}$. The direct part of the compression theorem is almost immediate from the definition of the sup-entropy rate in Eq. (7.11), so we briefly sketch it here: For $\delta > 0$ we define $\gamma := \overline{H}(\hat{\varrho}) + \delta$. As a compression projector we choose $P_n := \{\varrho_n \geq 2^{-n\gamma}\}$ as a substitute for the typical projector P_δ^n in the i.i.d. case discussed in Sec. 7.1.1. Since $\gamma > \overline{H}(\hat{\varrho})$, we may directly conclude from the definition in Eq. (7.11) that

$$\overline{\lim_{n\to\infty}}\ \mathrm{tr}\, P_n \varrho_n = 1\,, \tag{7.13}$$

and hence $\hat{\varrho}$ may be recovered faithfully. The use of the $\overline{\lim}$ in Eq. (7.13) as opposed to the lim in Eq. (7.7) means that we only require faithful compression along some (possibly sparse) subsequence. We could have made the same assumption for the i.i.d. case without changing the compression rate. This will be shown in Sec. 7.3.3 and is one of the fine point that will be swept under the rug in this preview.

The dimension of the spectral projection can now be estimated as follows:

$$1 \geq \operatorname{tr}\left\{\varrho_n \geq 2^{-n\gamma}\right\}\varrho_n = \sum_{\lambda_i^n \geq 2^{-n\gamma}} \lambda_i^n \;\geq 2^{-n\gamma} \operatorname{tr}\left\{\varrho_n \geq 2^{-n\gamma}\right\}, \tag{7.14}$$

and thus we find a compression rate

$$r = \frac{1}{n} \operatorname{ld} \operatorname{tr}\left\{\varrho_n \geq 2^{-n\gamma}\right\} \leq \gamma = \overline{H}(\hat{\varrho}) + \delta\,. \tag{7.15}$$

Since $\delta > 0$ is arbitrary, Eq. (7.15) shows that we may indeed compress the quantum source $\hat{\varrho}$ down to the sup-entropy rate $\overline{H}(\hat{\varrho})$, as suggested. In Sec. 7.3.1 we will prove that there is no better compression scheme. Hence, we have found an information-spectral expression for the optimal compression rate of a completely general quantum source $\hat{\varrho}$. If the source is memoryless, $\varrho_n = \varrho^{\otimes n}$ for some quantum state $\varrho \in \mathcal{B}_*(\mathcal{H})$, it is not difficult to show (cf. Sec. 7.3.3) that the sup-entropy rate indeed coincides with the standard von Neumann entropy, $\overline{H}(\hat{\varrho}) = H(\varrho) = -\operatorname{tr} \varrho \operatorname{ld} \varrho$.

7.1.3 Information Spectrum Methods

Information spectrum methods were conceived by Han and Verdú [HV93, VH94, Han03] as a novel coding technique for general sources and channels in classical information theory. As illustrated in Sec. 7.1.2, the underlying idea is to introduce spectral projections that interpolate between operational and more accessible information-theoretic quantities,

$$OP = SPEC = INF\,, \tag{7.16}$$

where $SPEC$ denotes information-spectrum quantities such as the sup-entropy rate in Eq. (7.11). The proof of the equality $OP = INF$ can then be broken into two parts: the proof $OP = SPEC$, which uses information-spectrum techniques and already contains all the essential coding arguments, and the proof $SPEC = INF$, which generally relies on properties of the spectrum of positive operators and basic probability theory.

Starting from a data compression theorem for general classical information sources [HV93], information-spectrum methods have been applied to a large variety of coding problems in classical information theory [Han03]. These ideas have been generalized to the quantum domain by Hayashi and Nagaoka for quantum hypothesis testing [NH02] and the classical capacity of a general quantum channel [HN03]. Hayashi [Hay03, Hay06^a] has also applied these ideas to derive a general formula for fixed-length entanglement concentration, the pure state variant of entanglement distillation.

In the remainder of this Chapter we will mostly be concerned with the data compression theorem for arbitrary quantum information sources. However, we will first make a slight detour via the quantum spectral divergence rates and some of their basic properties.

7.2 Quantum Spectral Divergence Rates

Quantum spectral divergence rates can be seen as a generalization of the quantum relative entropy (cf. App. A) and are the cornerstones of the information spectrum techniques. In fact, essentially all known coding theorems in classical and quantum information theory can be expressed in terms of divergence rates — albeit not usually in their most transparent form. The concept is hence much more general than what would be required to understand the proof of the data compression theorem in Sec. 7.3. Yet we take these extra steps to explain how data compression fits into the general information spectrum framework and to make better contact with the research literature.

7.2.1 Basic Definitions

The information spectrum approach requires the extensive use of spectral projections. Recall from Sec. 7.1.2 that for a self-adjoint operator $A \in \mathcal{B}(\mathcal{H})$ with spectral decomposition $A = \sum_i \lambda_i \, |i\rangle\langle i|$, we define the positive spectral projection on A as

$$\{A \geq 0\} := \sum_{\lambda_i \geq 0} |i\rangle\langle i| \,, \tag{7.17}$$

i. e., the projection onto the eigenspace of non-negative eigenvalues of A. Corresponding definitions apply for the other spectral projections $\{A > 0\}$, $\{A < 0\}$, and $\{A \leq 0\}$. For two self-adjoint operators $A, B \in \mathcal{B}(\mathcal{H})$, we can define $\{A \geq B\} := \{A - B \geq 0\}$, and similarly for the other three ordering relations. The quantum spectral divergence rates are defined in terms of these spectral projections as follows:

Definition 7.1. (Quantum Spectral Divergence Rates)
Let $(H_n)_{n\in\mathbb{N}}$ be a sequence of Hilbert spaces, and let $\hat{\varrho} := (\varrho_n)_{n\in\mathbb{N}}$ be a sequence of quantum states with $\varrho_n \in \mathcal{B}_*(\mathcal{H}_n)$. Assume that $\hat{\omega} := (\omega_n)_{n\in\mathbb{N}}$ is a sequence of positive operators (not necessarily normalized) with $\omega_n \in \mathcal{B}_*(\mathcal{H}_n)$. The quantum sup-divergence rate *of $\hat{\varrho}$ with respect to $\hat{\omega}$ is then defined as*

$$\overline{D}(\hat{\varrho} \,\|\, \hat{\omega}) := \inf \left\{ \gamma \;\middle|\; \overline{\lim_{n\to\infty}} \operatorname{tr} \{\varrho_n \geq 2^{n\gamma} \, \omega_n\} (\varrho_n - 2^{n\gamma} \, \omega_n) = 0 \right\}. \tag{7.18}$$

The quantum inf-divergence rate *of $\hat{\varrho}$ with respect to $\hat{\omega}$ is defined similarly as*

$$\underline{D}(\hat{\varrho} \,\|\, \hat{\omega}) := \sup \left\{ \gamma \;\middle|\; \overline{\lim_{n\to\infty}} \operatorname{tr} \{\varrho_n \geq 2^{n\gamma} \, \omega_n\} (\varrho_n - 2^{n\gamma} \, \omega_n) = 1 \right\}. \tag{7.19}$$

Before we explore basic properties of the spectral divergence rates in Sec. 7.2.3 we need to present two simple auxiliary lemmas.

7.2.2 Two Useful Lemmas

The proof of many of the basic properties of the spectral divergence rates relies on the following two simple lemmas, which we cite here from [BD06] and which we will employ frequently throughout the remainder of this Chapter.

Lemma 7.2. *Let $A, B \in \mathcal{B}(\mathcal{H})$ be two self-adjoint operators on some Hilbert space $\mathcal{H}$. Then for any positive operator $0 \leq P \leq \mathbb{1}$ we have*

$$\operatorname{tr} P\,(A-B) \leq \operatorname{tr} \{A \geq B\}(A-B)\,. \tag{7.20}$$

Proof: As both A and B are self-adjoint, so is $A-B$. Hence we can diagonalize $A-B$ and write it as the difference of two positive diagonal operators, $A-B = R-S$. Since the spectral projection $\{A \geq B\}$ projects onto the positive eigenvalues of $A-B$, we find for any positive operator $0 \leq P \leq \mathbb{1}$,

$$\operatorname{tr} \{A \geq B\}(A-B) = \operatorname{tr} R \geq \operatorname{tr}(P\,R) - \operatorname{tr}(P\,S) = \operatorname{tr} P\,(R-S) = \operatorname{tr} P\,(A-B)\,, \tag{7.21}$$

which is already the desired result. ■

Lemma 7.3. *For any state $\varrho \in \mathcal{B}_*(\mathcal{H})$ and any positive operator $\omega \in \mathcal{B}(\mathcal{H})$ we have*

$$\operatorname{tr} \{\varrho \geq c\,\omega\}\,\omega \leq \frac{1}{c} \tag{7.22}$$

for any $c > 0$.

Proof: Obviously, we have

$$\operatorname{tr} \{\varrho \geq c\,\omega\}(\varrho - c\,\omega) \geq 0\,. \tag{7.23}$$

Since ϱ is a quantum state, we also have $\operatorname{tr}(\{\varrho \geq c\,\omega\}\varrho) \leq 1$. The desired result then immediately follows from Eq. (7.23) by rearranging terms:

$$\operatorname{tr} \{\varrho \geq c\,\omega\}\,\omega \leq \frac{1}{c} \operatorname{tr} \{\varrho \geq c\,\omega\}\varrho \leq \frac{1}{c}\,. \quad \blacksquare \tag{7.24}$$

7.2.3 Basic Properties of the Quantum Spectral Divergence Rates

The quantum spectral divergence rates share many properties of the quantum relative entropy. In particular, Bowen and Datta have shown [BD06] that monotonicity under quantum operations and variants of chain rules and subadditivity relations all continue to hold for these asymptotic quantities. In this Section we will concentrate on those properties that we will later need for the proof of the compression theorem in Sec. 7.3.

Proposition 7.4. *Let $\hat{\varrho}$, $\hat{\omega}$ and the spectral divergence rates $\overline{D}(\hat{\varrho} \parallel \hat{\omega})$ and $\underline{D}(\hat{\varrho} \parallel \hat{\omega})$ be defined as in Def. 7.1 above. We then have,*

$$\underline{D}(\hat{\varrho} \parallel \hat{\omega}) \leq \overline{D}(\hat{\varrho} \parallel \hat{\omega})\,. \tag{7.25}$$

Proof: Let $\beta \in \mathbb{R}$ such that

$$\overline{\lim_{n\to\infty}} \operatorname{tr}\{\varrho_n \geq 2^{n\beta}\,\omega_n\}(\varrho_n - 2^{n\beta}\,\omega_n) = 0\,. \tag{7.26}$$

Lemma 7.2 and Lemma 7.3 then imply that for all $\gamma = \beta + \delta$ with some $\delta > 0$,

$$\begin{aligned} \operatorname{tr}\{\varrho_n \geq 2^{n\gamma}\omega_n\}\,\varrho_n &= \operatorname{tr}(\{\varrho_n \geq 2^{n\gamma}\,\omega_n\}\,(\varrho_n - 2^{n\beta}\,\omega_n)) + 2^{n\beta}\operatorname{tr}(\{\varrho_n \geq 2^{n\gamma}\omega_n\}\,\omega_n) \\ &\leq \operatorname{tr}(\{\varrho_n \geq 2^{n\beta}\,\omega_n\}\,(\varrho_n - 2^{n\beta}\,\omega_n)) + 2^{-n\delta}\,. \end{aligned} \tag{7.27}$$

We see from Eq. (7.26) that the RHS of Eq. (7.27) vanishes asymptotically as $n \to \infty$, implying that

$$\overline{\lim_{n\to\infty}} \operatorname{tr}\{\varrho_n \geq 2^{n\gamma}\,\omega_n\}(\varrho_n - 2^{n\gamma}\,\omega_n) = 0 \tag{7.28}$$

for every $\gamma = \beta + \delta$, and hence $\underline{D}(\hat{\varrho} \parallel \hat{\omega}) \leq \beta$, as suggested. ■

We will now show that the quantum spectral divergence rates separate high from low probability subspaces, as expected:

Proposition 7.5. *Let $\hat{\varrho}$, $\hat{\omega}$ and the spectral divergence rates $\overline{D}(\hat{\varrho} \parallel \hat{\omega})$ and $\underline{D}(\hat{\varrho} \parallel \hat{\omega})$ be defined as in Def. 7.1 above. Then for any $\gamma > \overline{D}(\hat{\varrho} \parallel \hat{\omega})$ we have*

$$\overline{\lim_{n\to\infty}} \operatorname{tr}\{\varrho_n \geq 2^{n\gamma}\,\omega_n\}(\varrho_n - 2^{n\gamma}\,\omega_n) = 0\,. \tag{7.29}$$

Similarly, for any $\gamma < \underline{D}(\hat{\varrho} \parallel \hat{\omega})$ we have

$$\overline{\lim_{n\to\infty}} \operatorname{tr}\{\varrho_n \geq 2^{n\gamma}\,\omega_n\}(\varrho_n - 2^{n\gamma}\,\omega_n) = 1\,. \tag{7.30}$$

Proof: Assuming $\gamma > \overline{D}(\hat{\varrho} \parallel \hat{\omega})$, we may find some $\beta \in [\overline{D}(\hat{\varrho} \parallel \hat{\omega}), \gamma)$ such that

$$\overline{\lim_{n\to\infty}} \operatorname{tr}\{\varrho_n \geq 2^{n\beta}\,\omega_n\}(\varrho_n - 2^{n\beta}\,\omega_n) = 0\,. \tag{7.31}$$

Setting $\delta := \gamma - \beta > 0$, we may then apply Lemma 7.2 and Lemma 7.3 to conclude that

$$\begin{aligned} \operatorname{tr}\{\varrho_n \geq 2^{n\gamma}\,\omega_n\}(\varrho_n - 2^{n\gamma}\,\omega_n) &\leq \operatorname{tr}\{\varrho_n \geq 2^{n\gamma}\,\omega_n\}\varrho_n \\ &= \operatorname{tr}\{\varrho_n \geq 2^{n\gamma}\,\omega_n\}(\varrho_n - 2^{n\beta}\,\omega_n) + 2^{n\beta}\operatorname{tr}\{\varrho_n \geq 2^{n\gamma}\,\omega_n\}\,\omega_n \\ &\leq \operatorname{tr}\{\varrho_n \geq 2^{n\beta}\,\omega_n\}(\varrho_n - 2^{n\beta}\,\omega_n) + 2^{n\beta}\,2^{-n\gamma} \\ &\leq \operatorname{tr}\{\varrho_n \geq 2^{n\beta}\,\omega_n\}(\varrho_n - 2^{n\beta}\,\omega_n) + 2^{-n\delta}\,. \end{aligned} \tag{7.32}$$

Taking the limes superior on both sides of Eq. (7.32), we immediately see from Eq. (7.31) that

$$\overline{\lim_{n\to\infty}} \operatorname{tr}\{\varrho_n \geq 2^{n\gamma}\,\omega_n\}(\varrho_n - 2^{n\gamma}\,\omega_n) = 0\,, \tag{7.33}$$

as suggested. ▲

The proof of Eq. (7.30) is very much along the same lines: Since $\gamma < \underline{D}(\hat{\varrho} \parallel \hat{\omega})$, we may find some $\beta \in (\gamma, \underline{D}(\hat{\varrho} \parallel \hat{\omega})]$ such that

$$\overline{\lim_{n\to\infty}} \operatorname{tr} \{\varrho_n \geq 2^{n\beta}\, \omega_n\}(\varrho_n - 2^{n\beta}\, \omega_n) = 1\,. \tag{7.34}$$

Hence, setting $\delta := \beta - \gamma > 0$ we conclude from Eq. (7.32) (with the roles of γ and β interchanged) that

$$\operatorname{tr} \{\varrho_n \geq 2^{n\beta}\, \omega_n\}(\varrho_n - 2^{n\beta}\, \omega_n) \leq \operatorname{tr} \{\varrho_n \geq 2^{n\gamma}\, \omega_n\}(\varrho_n - 2^{n\gamma}\, \omega_n) + 2^{-n\delta} \leq 1 + 2^{-n\delta}\,. \tag{7.35}$$

The result then follows from Eq. (7.34) by letting $n \to \infty$. ■

Bowen and Datta have shown [BD06] that the spectral divergence rates may be given a slightly simplified form, which will frequently prove helpful in the remainder of this Chapter:

Proposition 7.6. *Let $\hat{\varrho}$, $\hat{\omega}$ and the spectral divergence rates $\overline{D}(\hat{\varrho} \parallel \hat{\omega})$ and $\underline{D}(\hat{\varrho} \parallel \hat{\omega})$ be defined as in Def. 7.1 above. We then have:*

$$\overline{D}(\hat{\varrho} \parallel \hat{\omega}) = \inf \left\{\gamma \;\middle|\; \overline{\lim_{n\to\infty}} \operatorname{tr} \{\varrho_n \geq 2^{n\gamma}\, \omega_n\}\, \varrho_n = 0\right\} =: \overline{D}'(\hat{\varrho} \parallel \hat{\omega}) \quad \text{and} \tag{7.36}$$

$$\underline{D}(\hat{\varrho} \parallel \hat{\omega}) = \sup \left\{\gamma \;\middle|\; \overline{\lim_{n\to\infty}} \operatorname{tr} \{\varrho_n \geq 2^{n\gamma}\, \omega_n\}\, \varrho_n = 1\right\} =: \underline{D}'(\hat{\varrho} \parallel \hat{\omega})\,. \tag{7.37}$$

A comparison with Eq. (7.11) then immediately shows that

$$\begin{aligned}
\overline{H}(\hat{\varrho}) = -\underline{D}(\hat{\varrho} \parallel \mathbb{1}) &= -\sup \left\{\gamma \;\middle|\; \overline{\lim_{n\to\infty}} \operatorname{tr} \{\varrho_n \geq 2^{n\gamma}\, \mathbb{1}_n\}(\varrho_n - 2^{n\gamma}\, \mathbb{1}_n) = 1\right\} \\
&= \inf \left\{\gamma \;\middle|\; \overline{\lim_{n\to\infty}} \operatorname{tr} \{\varrho_n \geq 2^{-n\gamma}\, \mathbb{1}_n\}\, (\varrho_n - 2^{-n\gamma}\, \mathbb{1}_n) = 1\right\} \\
&= \inf \left\{\gamma \;\middle|\; \overline{\lim_{n\to\infty}} \operatorname{tr} \{\varrho_n \geq 2^{-n\gamma}\, \mathbb{1}_n\}\, \varrho_n = 1\right\} \\
&= \inf \left\{\gamma \;\middle|\; \underline{\lim_{n\to\infty}} \operatorname{tr} \{\varrho_n < 2^{-n\gamma}\, \mathbb{1}_n\}\, \varrho_n = 0\right\},
\end{aligned} \tag{7.38}$$

where in the second to last step we have made use of Prop. 7.6, and in the last step we have used that $\{\varrho_n < 2^{-n\gamma}\, \mathbb{1}_n\} = \mathbb{1}_n - \{\varrho_n \geq 2^{-n\gamma}\, \mathbb{1}_n\}$ and that $\operatorname{tr} \varrho_n = 1$. Hence, we recover the sup-entropy rate as a special case of the quantum inf-divergence rate. The proof of Prop. 7.6 may be found in Sec. III.A of [BD06]. We reproduce it here for completeness.

Proof of Prop. 7.6: For any $\gamma \in \mathbb{R}$ we have,

$$1 \geq \operatorname{tr} \{\varrho_n \geq 2^{n\gamma}\}\, \varrho_n \geq \operatorname{tr} \{\varrho_n \geq 2^{n\gamma}\}\, (\varrho_n - 2^{n\gamma}) \geq 0\,. \tag{7.39}$$

Hence, if the LHS of Eq. (7.39) vanishes asymptotically, so does the RHS, implying that $\overline{D}(\hat{\varrho} \parallel \hat{\omega}) \leq \overline{D}'(\hat{\varrho} \parallel \hat{\omega})$ Correspondingly, if the RHS of Eq. (7.39) approaches 1 asymptotically, the same holds true for the LHS, and thus $\underline{D}(\hat{\varrho} \parallel \hat{\omega}) \leq \underline{D}'(\hat{\varrho} \parallel \hat{\omega})$.

For the converse implications we now choose $\delta > 0$ and set $\gamma := \overline{D}(\hat{\varrho} \parallel \hat{\omega}) + \delta$ and $\beta := \gamma + \delta = \overline{D}(\hat{\varrho} \parallel \hat{\omega}) + 2\,\delta$. Lemma 7.2 and Lemma 7.3 then immediately imply that

$$\begin{aligned} \operatorname{tr}\left\{\varrho_n \geq 2^{n\beta}\,\omega_n\right\}\varrho_n &= \operatorname{tr}\left\{\varrho_n \geq 2^{n\beta}\,\omega_n\right\}(\varrho_n - 2^{n\gamma}\,\omega_n) + 2^{n\gamma}\operatorname{tr}\left\{\varrho_n \geq 2^{n\beta}\,\omega_n\right\}\omega_n \\ &\leq \operatorname{tr}\left\{\varrho_n \geq 2^{n\gamma}\,\omega_n\right\}(\varrho_n - 2^{n\gamma}\,\omega_n) + 2^{-n\delta}\,. \end{aligned} \tag{7.40}$$

Taking the limes superior, we conclude from Prop. 7.5 that the RHS of Eq. (7.40) vanishes, because $\gamma > \overline{D}(\hat{\varrho} \parallel \hat{\omega})$. We thus have $\overline{D}'(\hat{\varrho} \parallel \hat{\omega}) \leq \beta = \overline{D}(\hat{\varrho} \parallel \hat{\omega}) + 2\,\delta$. Since this holds for arbitrary $\delta > 0$, we have $\overline{D}'(\hat{\varrho} \parallel \hat{\omega}) \leq \overline{D}(\hat{\varrho} \parallel \hat{\omega})$, the desired result.

The proof of the remaining bound, $\underline{D}'(\hat{\varrho} \parallel \hat{\omega}) \leq \underline{D}(\hat{\varrho} \parallel \hat{\omega})$, is again very similar. For $\delta > 0$, we set $\beta := \underline{D}'(\hat{\varrho} \parallel \hat{\omega}) - \delta$ and $\gamma := \beta - \delta = \underline{D}'(\hat{\varrho} \parallel \hat{\omega}) - 2\,\delta$. We then conclude that the LHS in Eq. (7.40) approaches 1 asymptotically, and hence

$$\overline{\lim_{n\to\infty}} \operatorname{tr}\left\{\varrho_n \geq 2^{n\gamma}\,\omega_n\right\}(\varrho_n - 2^{n\gamma}\,\omega_n) = 1\,, \tag{7.41}$$

implying that $\underline{D}(\hat{\varrho} \parallel \hat{\omega}) \geq \gamma = \underline{D}'(\hat{\varrho} \parallel \hat{\omega}) - 2\,\delta$. Since $\delta > 0$ is arbitrary, this concludes the proof. ■

7.3 Quantum Source Coding

The basic idea and philosophy behind quantum source coding has already been explained in quite some detail in Sec. 7.1.1. As illustrated in Fig. 7.1, a compression scheme of rate r for the general quantum source $\hat{\varrho} = (\varrho_n)_{n\in\mathbb{N}}$ with $\varrho_n \in \mathcal{B}_*(\mathcal{H}_n)$ consists of two families of quantum operations, $(C_n)_{n\in\mathbb{N}}$ and $(D_n)_{n\in\mathbb{N}}$. The compression channel C_n takes states in the d_n-dimensional Hilbert space $\mathcal{H}_n$ to states in a 2^{nr}-dimensional subspace. We can regard the compressed space as representing nr qubits. The quantum channel D_n is the decompression operation, which takes states in the compressed space to states in the original state space $\mathcal{H}_n$. Our criterion for reliability is that in the asymptotic limit the entanglement with a bystander system is preserved, such that $F_e(\varrho_n, D_{n*} \circ C_{n*}) \to 1$ as $n \to \infty$. Hence, we cast the following

Definition 7.7. (Quantum Data Compression)
Let $(\mathcal{H}_n)_{n\in\mathbb{N}}$ be a sequence of finite-dimensional Hilbert spaces, and let $\hat{\varrho} := (\varrho_n)_{n\in\mathbb{N}}$ with $\varrho_n \in \mathcal{B}_*(\mathcal{H}_n)$ be a quantum source. (The ϱ_n have to be density operators, but no further structural assumptions or consistency conditions are made.) A number $r > 0$ is called an achievable rate *for the quantum source $\hat{\varrho}$ iff there exists an integer sequence $(n_\nu)_{\nu\in\mathbb{N}}$ with $\lim_{\nu\to\infty} n_\nu = \infty$ and sequences of compression operations $C_{n_\nu *}\colon \mathcal{B}_*(\mathcal{H}_{n_\nu}) \to \mathcal{B}_*(\mathcal{K}_{n_\nu})$ and decompression operations $D_{n_\nu *}\colon \mathcal{B}_*(\mathcal{K}_{n_\nu}) \to \mathcal{B}_*(\mathcal{H}_{n_\nu})$ such that*

$$\lim_{\nu\to\infty} \frac{\operatorname{ld}\dim \mathcal{K}_{n_\nu}}{n_\nu} \leq r \qquad \text{and} \qquad \lim_{\nu\to\infty} F_e(\varrho_{n_\nu}, D_{n_\nu *} \circ C_{n_\nu *}) = 1\,, \tag{7.42}$$

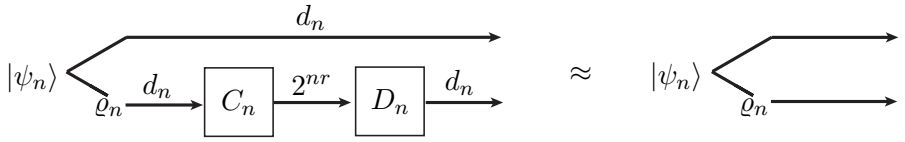

Figure 7.1: *A data compression scheme for the quantum source $\hat{\varrho} = (\varrho_n)_{n\in\mathbb{N}}$ with compression operations C_n, decompression operations D_n, and compression rate r. In the asymptotic limit $n \to \infty$, the entanglement with a bystander system is faithfully preserved: if $|\psi_n\rangle$ is a purification of ϱ_n, we require* $(\mathrm{id}_n \otimes D_{n*} \circ C_{n*})|\psi_n\rangle\langle\psi_n| \approx |\psi_n\rangle\langle\psi_n|$.

where F_e denotes the entanglement fidelity*, as introduced in Sec. 5.2.2. The* compression rate *$R(\hat{\varrho})$ of the quantum source $\hat{\varrho}$ is defined as the infimum of all achievable rates.*

Remark 7.8. In Def. 7.7 we have only required that there exists *one* sequence of compression spaces along which the rate r is achieved and faithful decompression is feasible. A more ambitious definition would require that the entanglement fidelity approaches 1 along *any* sub-sequence $(n_\nu)_{\nu\in\mathbb{N}}$ such that $\lim_{\nu\to\infty} \frac{\mathrm{ld}\,\dim \mathcal{K}_{n_\nu}}{n_\nu} \geq r$ — similar to our definition of quantum channel capacity in Sec. 5.2. These different capacity concepts are sometimes called the "optimistic" and the "pessimistic" point of view (cf. [CK81], p. 128). For general sources $\hat{\varrho}$, these viewpoints may lead to different values of the compression rate. However, we will show in Sec. 7.3.3 below that they coincide when $\hat{\varrho}$ is ergodic — just as they do for memoryless (cf. Sec. 5.2.1) or forgetful quantum channels (cf. Remark 6.25). As explained in Sec. 7.3.2, ergodic sources even satisfy a *strong converse:* we cannot achieve smaller compression rates by allowing finite compression errors.

Remark 7.9. Note that the compression scheme in Def. 7.7 is allowed to depend on the quantum source $\hat{\varrho}$. *Universal* compression schemes for i.i.d. quantum sources have been designed by Jozsa *et al.* [JHH$^+$98], and more generally for ergodic sources by Kaltchenko and Yang [KY03]. We will come back to these results when we discuss mixed quantum sources in Sec. 7.3.4.

7.3.1 A Data Compression Theorem for General Quantum Sources

As advertised in Sec. 7.1.2, we will now apply the quantum information spectrum methods to prove a source coding theorem for general quantum data sources. A strong converse will then be shown in Sec. 7.3.2. In Sec. 7.3.3 we will apply our results to ergodic quantum sources, and in Sec. 7.3.4 to mixed quantum sources.

Theorem 7.10. (Source Coding for General Quantum Sources)
Let $(\mathcal{H}_n)_{n\in\mathbb{N}}$ be a sequence of finite-dimensional Hilbert spaces, and let $\hat{\varrho} := (\varrho_n)_{n\in\mathbb{N}}$

with $\varrho_n \in \mathcal{B}_*(\mathcal{H}_n)$ be a quantum source with compression rate $R(\hat{\varrho})$, as defined in Def. 7.7 above. We then have

$$R(\hat{\varrho}) = \overline{H}(\hat{\varrho})\,, \tag{7.43}$$

where $\overline{H}(\hat{\varrho}) = -\underline{D}(\hat{\varrho} \parallel \mathbb{1})$ denotes the spectral sup-entropy rate *of $\hat{\varrho}$, as defined in Eq. (7.11).*

Proof: We start with the direct part. For $r > \overline{H}(\hat{\varrho})$ and a quantum state $\sigma \in \mathcal{B}_*(\mathcal{H}_n)$ we define the compression operation

$$C_{n*}(\sigma) := P_n \sigma P_n + \mathrm{tr}((\mathbb{1}_n - P_n)\,\sigma)\,|\varphi_n\rangle\langle\varphi_n|\,, \tag{7.44}$$

where we have set $P_n := \{\varrho_n \geq 2^{-nr}\mathbb{1}_n\}$, and φ_n denotes an arbitrary pure state on the projected space $P_n\mathcal{H}_n \subset \mathcal{H}_n$, so that C_{n*} is both completely positive and trace-preserving. The corresponding decompression operation will then be chosen simply as the identity channel on $P_n\mathcal{H}_n$, $D_n = \mathrm{id}_n$. From Eq. (5.13) we find

$$1 \geq F_e(\varrho_n, D_{n*}\,C_{n*}) \geq |\mathrm{tr}\; P_n\,\varrho_n|^2 = |\mathrm{tr}\;\{\varrho_n \geq 2^{-nr}\mathbb{1}_n\}\,\varrho_n|^2\,. \tag{7.45}$$

Since $r > \overline{H}(\hat{\varrho})$, we can now immediately conclude from the definition of the spectral sup-entropy rate $\overline{H}(\hat{\varrho}) = -\underline{D}(\hat{\varrho} \parallel \mathbb{1})$ that there exists a subsequence $(n_\nu)_{\nu\in\mathbb{N}}$ such that

$$\lim_{\nu\to\infty} F_e(\varrho_{n_\nu}, D_{n_\nu *} \circ C_{n_\nu *}) = 1\,, \tag{7.46}$$

and hence this compression scheme is reliable. From the relation

$$1 \geq \mathrm{tr}\;\{\varrho_{n_\nu} \geq 2^{-n_\nu r}\mathbb{1}_{n_\nu}\}\,\varrho_{n_\nu} = \sum_{\lambda_i^{n_\nu} \geq 2^{-n_\nu r}} \lambda_i^{n_\nu} \geq 2^{-n_\nu r}\,\mathrm{tr}\;\{\varrho_{n_\nu} \geq 2^{-n_\nu r}\mathbb{1}_{n_\nu}\} \tag{7.47}$$

we find that the compressed space $P_n\mathcal{H}_n \subset \mathcal{H}_n$ has dimension

$$\mathrm{tr}\; P_{n_\nu} = \mathrm{tr}\;\{\varrho_{n_\nu} \geq 2^{-n_\nu r}\mathbb{1}_{n_\nu}\} \leq 2^{n_\nu r}\,, \tag{7.48}$$

and hence r is an achievable rate according to Def. 7.7. Since $r > \overline{H}(\hat{\varrho})$ is arbitrary, we conclude that $R(\hat{\varrho}) \leq \overline{H}(\hat{\varrho})$, as suggested. ▲

For the converse implication, assume that $r < \overline{H}(\hat{\varrho})$ and set $\overline{H}(\hat{\varrho}) - r =: 2\,\delta$ for some $\delta > 0$. Assume that we have some subsequence $(n_\nu)_{\nu\in\mathbb{N}}$ and a sequence of Hilbert spaces $(\mathcal{K}_{n_\nu})_{\nu\in\mathbb{N}}$ with $\mathrm{ld}\,\dim\mathcal{K}_{n_\nu} \leq n_\nu\,r$. A corresponding coding scheme consists of compression operations $C_{n_\nu *}\colon \mathcal{B}_*(\mathcal{H}_{n_\nu}) \to \mathcal{B}_*(\mathcal{K}_{n_\nu})$ and decompression operations $D_{n_\nu *}\colon \mathcal{B}_*(\mathcal{K}_{n_\nu}) \to \mathcal{B}_*(\mathcal{H}_{n_\nu})$. Let $\{c_{n_\nu}^k\}_k$ and $\{d_{n_\nu}^j\}_j$ be two sets of Kraus operators for the quantum channels C_{n_ν} and D_{n_ν}, respectively, and let P_{n_ν} denote the projection on the subspace $\mathcal{K}_{n_\nu} \subset \mathcal{H}_{n_\nu}$. If $P_{n_\nu}^j$ denotes the projector onto the subspace of $\mathcal{H}_n$ to which P_{n_ν} is mapped under the Kraus operator $d_{n_\nu}^j$, we have $d_{n_\nu}^j c_{n_\nu}^k = d_{n_\nu}^j P_{n_\nu} c_{n_\nu}^k = P_{n_\nu}^j d_{n_\nu}^j c_{n_\nu}^k$.

Applying Eq. (5.13) and the Cauchy-Schwarz inequality for the Hilbert-Schmidt inner product, the entanglement fidelity may now be bounded as follows:

$$\begin{aligned} F_e(\varrho_{n_\nu}, D_{n_\nu *} \circ C_{n_\nu *}) &= \sum_{j,k} |\mathrm{tr}\; d^j_{n_\nu} c^k_{n_\nu} \varrho_{n_\nu}|^2 \\ &= \sum_{j,k} |\mathrm{tr}\; P^j_{n_\nu} d^j_{n_\nu} c^k_{n_\nu} \varrho_{n_\nu}|^2 \\ &\leq \sum_{j,k} \mathrm{tr}(P^j_{n_\nu} \varrho_{n_\nu} P^j_{n_\nu})\, \mathrm{tr}(d^j_{n_\nu} c^k_{n_\nu} \varrho_{n_\nu} c^{k*}_{n_\nu} d^{j*}_{n_\nu}) \\ &\leq \mathrm{tr}(P_{n_\nu} \varrho_{n_\nu} P_{n_\nu}) \sum_{j,k} \mathrm{tr}\; d^j_{n_\nu} c^k_{n_\nu} \varrho_{n_\nu} c^{k*}_{n_\nu} d^{j*}_{n_\nu} \\ &= \mathrm{tr}(P_{n_\nu}(\varrho_{n_\nu} - 2^{-n_\nu \gamma}\, \mathbb{1}_{n_\nu})) + 2^{-n_\nu \gamma} \mathrm{tr}\; P_{n_\nu} \\ &\leq \mathrm{tr}(\{\varrho_{n_\nu} \geq 2^{-n_\nu \gamma}\, \mathbb{1}_{n_\nu}\}(\varrho_{n_\nu} - 2^{-n_\nu \gamma}\, \mathbb{1}_{n_\nu})) + 2^{-n_\nu \gamma} \mathrm{tr}\; P_{n_\nu} \end{aligned} \tag{7.49}$$

for any $\gamma \in \mathbb{R}$, where in the last step we have again made use of Lemma 7.2. Choosing $\gamma := \overline{H}(\hat{\varrho}) - \delta = r + \delta$ and using that $\mathrm{tr}\; P_{n_\nu} = \dim \mathcal{K}_{n_\nu} \leq 2^{n_\nu r}$, we now conclude from Eq. (7.49) that

$$F_e(\varrho_{n_\nu}, D_{n_\nu *} \circ C_{n_\nu *}) \leq \mathrm{tr}\left(\{\varrho_{n_\nu} \geq 2^{-n_\nu \gamma}\, \mathbb{1}_{n_\nu}\}(\varrho_{n_\nu} - 2^{-n_\nu \gamma}\, \mathbb{1}_{n_\nu})\right) + 2^{-n_\nu \delta}\,. \tag{7.50}$$

Since $\gamma < \overline{H}(\hat{\varrho})$, the first term on the RHS of Eq. (7.50) is bounded away from 1, and so is $F_e(\varrho_{n_\nu}, D_{n_\nu *} \circ C_{n_\nu *})$, as the second term vanishes as $\nu \to \infty$. Hence, no such compression scheme with rate $r < \overline{H}(\hat{\varrho})$ can be reliable. ■

7.3.2 Strong Converse

In the proof of Th. 7.10 we have shown a *weak converse* for quantum data compression: if we try to compress a quantum source $\hat{\varrho}$ at a rate below $\overline{H}(\hat{\varrho})$, then the fidelity will be bounded away from 1, and hence we will have to live with finite errors. The information-spectrum method can also be applied to formulate a *strong converse* in terms of the so-called *spectral inf-entropy* rate,

$$\begin{aligned} \underline{H}(\hat{\varrho}) := -\overline{D}(\hat{\varrho} \parallel \mathbb{1}) &= -\inf\left\{\gamma \;\middle|\; \overline{\lim_{n\to\infty}}\, \mathrm{tr}\,\{\varrho_n \geq 2^{n\gamma}\, \mathbb{1}_n\}(\varrho_n - 2^{n\gamma}\, \mathbb{1}_n) = 0\right\} \\ &= \sup\left\{\gamma \;\middle|\; \overline{\lim_{n\to\infty}}\, \mathrm{tr}\,\{\varrho_n \geq 2^{-n\gamma}\, \mathbb{1}_n\}\,(\varrho_n - 2^{-n\gamma}\, \mathbb{1}_n) = 0\right\} \\ &= \sup\left\{\gamma \;\middle|\; \overline{\lim_{n\to\infty}}\, \mathrm{tr}\,\{\varrho_n \geq 2^{-n\gamma}\, \mathbb{1}_n\}\,\varrho_n = 0\right\} \\ &= \sup\left\{\gamma \;\middle|\; \underline{\lim_{n\to\infty}}\, \mathrm{tr}\,\{\varrho_n < 2^{-n\gamma}\, \mathbb{1}_n\}\,\varrho_n = 1\right\}, \end{aligned} \tag{7.51}$$

where in the second to last step we have again made use of Prop. 7.6, and in the last step we have used that $\{\varrho_n < 2^{-n\gamma}\, \mathbb{1}_n\} = \mathbb{1}_n - \{\varrho_n \geq 2^{-n\gamma}\, \mathbb{1}_n\}$ and that $\mathrm{tr}\;\varrho_n = 1$. Whenever we try to compress the source $\hat{\varrho}$ at rates below $\underline{H}(\hat{\varrho})$, then the error probability will approach 1, and hence we cannot get away with small finite errors:

Theorem 7.11. (Strong Converse)
Let $(\mathcal{H}_n)_{n\in\mathbb{N}}$ be a sequence of finite-dimensional Hilbert spaces, and let $\hat{\varrho} := (\varrho_n)_{n\in\mathbb{N}}$ with $\varrho_n \in \mathcal{B}_*(\mathcal{H}_n)$ be a quantum source. Assume a compression scheme with compression operators $(C_{n_\nu})_{\nu\in\mathbb{N}}$ and decompression operators $(D_{n_\nu})_{\nu\in\mathbb{N}}$ with rate $r < \underline{H}(\hat{\varrho})$. We then have

$$\lim_{\nu\to\infty} F_e(\varrho_{n_\nu}, D_{n_\nu *} \circ C_{n_\nu *}) = 0\,. \tag{7.52}$$

Proof: For $r < \underline{H}(\hat{\varrho})$ we set $\underline{H}(\hat{\varrho}) - r =: 2\,\delta$ for some $\delta > 0$ and $\gamma := \underline{H}(\hat{\varrho}) - \delta$. We can then conclude from the definition of the spectral inf-entropy rate that

$$\lim_{\nu\to\infty} \mathrm{tr}\,\{\varrho_{n_\nu} \geq 2^{-n_\nu\gamma}\,\mathbb{1}_{n_\nu}\}\,(\varrho_{n_\nu} - 2^{-n_\nu\gamma}\,\mathbb{1}_{n_\nu}) = 0\,, \tag{7.53}$$

and the desired result immediately follows from Eq. (7.50). ■

A quantum source $\hat{\varrho}$ such that $\underline{H}(\hat{\varrho}) = \overline{H}(\hat{\varrho})$ is usually called *information-stable*. For such sources, there is no need to distinguish a weak and a strong converse: we cannot achieve better compression rates by tolerating small finite errors. We will show in the next Section that ergodic sources are information-stable. Examples of sources where $\underline{H}(\hat{\varrho}) \neq \overline{H}(\hat{\varrho})$ will then be presented in Sec. 7.3.4.

7.3.3 Ergodic Sources

We start by showing how the spectral entropy rates are related to the von Neumann entropy:

Proposition 7.12. *Let $(\mathcal{H}_n)_{n\in\mathbb{N}}$ be a sequence of finite-dimensional Hilbert spaces, and let $\hat{\varrho} := (\varrho_n)_{n\in\mathbb{N}}$ with $\varrho_n \in \mathcal{B}_*(\mathcal{H}_n)$ be a quantum source. We then have:*

$$\underline{H}(\hat{\varrho}) \leq \underline{\lim_{n\to\infty}} \frac{1}{n} H(\varrho_n) \leq \overline{\lim_{n\to\infty}} \frac{1}{n} H(\varrho_n) \leq \overline{H}(\hat{\varrho})\,, \tag{7.54}$$

where $H(\sigma) = -\mathrm{tr}\,\sigma\,\mathrm{ld}\,\sigma$ denotes the von Neumann entropy.

Proof: Let $\{\lambda_i^n\}_i$ denote the set of eigenvalues of the density operator $\varrho_n \in \mathcal{B}_*(\mathcal{H}_n)$. For $\delta > 0$ we then introduce the shorthand $\gamma := \underline{H}(\hat{\varrho}) - \delta$ and estimate

$$\frac{1}{n} H(\varrho_n) = -\frac{1}{n} \sum_i \lambda_i^n\,\mathrm{ld}\,\lambda_i^n \geq -\frac{1}{n} \sum_{\lambda_i^n < 2^{-n\gamma}} \lambda_i^n\,\mathrm{ld}\,\lambda_i^n \geq \gamma\,\mathrm{tr}\,\{\varrho_n < 2^{-n\gamma}\mathbb{1}_n\}\,\varrho_n\,. \tag{7.55}$$

Since $\gamma < \underline{H}(\hat{\varrho})$, we find from the definition of the spectral inf-entropy rate in Eq. (7.51) that

$$\underline{\lim_{n\to\infty}} \mathrm{tr}\,\{\varrho_n < 2^{-n\gamma}\mathbb{1}_n\}\,\varrho_n = 1\,. \tag{7.56}$$

Since this holds true for all $\delta > 0$, we then immediately conclude from Eq. (7.55) that

$$\underline{H}(\hat{\varrho}) \leq \underline{\lim_{n\to\infty}} \frac{1}{n} H(\varrho_n)\,. \quad ▲ \tag{7.57}$$

For $\delta > 0$ we now set $\beta := \overline{H}(\hat{\varrho}) + \delta$ and define the projector $Q_n := \{\varrho_n < 2^{-n\beta}\, \mathbb{1}_n\}$. We thus have,

$$\begin{aligned}\frac{1}{n} H(\varrho_n) &= -\frac{1}{n} \sum_i \lambda_i^n \operatorname{ld} \lambda_i^n \\ &= -\frac{1}{n} \sum_{\lambda_i^n \geq 2^{-n\beta}} \lambda_i^n \operatorname{ld} \lambda_i^n \; - \frac{1}{n} \sum_{\lambda_i^n < 2^{-n\beta}} \lambda_i^n \operatorname{ld} \lambda_i^n \\ &\leq \beta \operatorname{tr} \{\varrho_n \geq 2^{-n\beta}\, \mathbb{1}_n\}\, \varrho_n \; - \; \frac{1}{n} \operatorname{tr}\, Q_n \varrho_n \operatorname{ld} \varrho_n \,. \end{aligned} \tag{7.58}$$

Using the operator monotonicity of the logarithm [Löw34, Bha97], we have

$$\begin{aligned}\operatorname{tr}\, Q_n\, \varrho_n\, (-\operatorname{ld} \varrho_n) &\leq \operatorname{tr}\, Q_n\, \varrho_n\, (-\operatorname{ld} Q_n \varrho_n Q_n) \\ &= \operatorname{tr}\, Q_n\, \varrho_n\, (-\operatorname{ld} Q_n^2 \varrho_n Q_n^2) \\ &= \operatorname{tr}\, Q_n^2\, \varrho_n\, Q_n\, (-\operatorname{ld} Q_n \varrho_n Q_n) \\ &= -\operatorname{tr}\, \tilde{\sigma}_n \operatorname{ld} \tilde{\sigma}_n \,, \end{aligned} \tag{7.59}$$

where we have set $\tilde{\sigma}_n := Q_n \varrho_n Q_n$. Introducing the normalized state $\sigma_n := \operatorname{tr}^{-1}(\tilde{\sigma}_n)\, \tilde{\sigma}_n$, we then conclude from Eqs. (7.58) and (7.59) that

$$\begin{aligned}\frac{1}{n} H(\varrho_n) &\leq \beta \operatorname{tr} \{\varrho_n \geq 2^{-n\beta}\, \mathbb{1}_n\}\, \varrho_n - \frac{1}{n} \operatorname{tr}\, \tilde{\sigma}_n (\operatorname{ld} \tilde{\sigma}_n - \operatorname{ld} \operatorname{tr}\, \tilde{\sigma}_n + \operatorname{ld} \operatorname{tr}\, \tilde{\sigma}_n) \\ &\leq \beta \operatorname{tr} \{\varrho_n \geq 2^{-n\beta}\, \mathbb{1}_n\}\, \varrho_n + \frac{1}{n} \operatorname{tr}(\tilde{\sigma}_n)\, H(\sigma_n) - \frac{1}{n} \operatorname{tr}(\tilde{\sigma}_n) \operatorname{ld} \operatorname{tr}(\tilde{\sigma}_n) \\ &\leq \beta \operatorname{tr} \{\varrho_n \geq 2^{-n\beta}\, \mathbb{1}_n\}\, \varrho_n + \frac{1}{n} \operatorname{tr}(\tilde{\sigma}_n) \operatorname{ld} d_n - \frac{1}{n} \operatorname{tr}(\tilde{\sigma}_n) \operatorname{ld} \operatorname{tr}(\tilde{\sigma}_n) \,, \end{aligned} \tag{7.60}$$

where we have set $d_n := \dim \mathcal{H}_n$. In the following we assume that there exists some $\alpha > 0$ such that $\frac{1}{n} \operatorname{ld} d_n \leq \alpha$ for all $n \in \mathbb{N}$. Since $\beta > \overline{H}(\hat{\varrho})$, we then have

$$\underline{\lim}_{n\to\infty} \operatorname{tr}\, \tilde{\sigma}_n = \underline{\lim}_{n\to\infty} \operatorname{tr}\, \{\varrho_n < 2^{-n\beta}\, \mathbb{1}_n\}\, \varrho_n = 0 \tag{7.61}$$

from Eq. (7.38). The second and the third term on the RHS of Eq. (7.60) hence vanish in the asymptotic limit, and we conclude that

$$\overline{\lim_{n\to\infty}} \frac{1}{n} H(\varrho_n) \leq \beta = \overline{H}(\hat{\varrho}) + \delta \,. \tag{7.62}$$

Since $\delta > 0$ is arbitrary, we have $\overline{\lim}_{n\to\infty} \frac{1}{n} H(\varrho_n) \leq \overline{H}(\hat{\varrho})$. The remaining inequality in Eq. (7.54) is clear from the definition of $\underline{\lim}$ and $\overline{\lim}$. ■

If the quantum source $\hat{\varrho}$ is ergodic, we can apply the quantum ergodic equipartition theorem [BKS$^+$04, BKS$^+$03] to show that all the four quantities in Eq. (7.54) coincide. This result is originally due to Bjelaković and Szkoła [BS05]. We include it here as an illustrative example of how to reduce information-spectrum quantities to more standard entropic expressions.

Proposition 7.13. (Data Compression for Ergodic Sources)
Let $(\mathcal{H}_n)_{n\in\mathbb{N}}$ be a sequence of finite-dimensional Hilbert spaces, and let $\hat{\varrho} := (\varrho_n)_{n\in\mathbb{N}}$ with $\varrho_n \in \mathcal{B}_*(\mathcal{H}_n)$ be an ergodic quantum source, i.e., a source which is extremal in the set of translational invariant sources (cf. Appendix C). We then have,

$$\underline{H}(\hat{\varrho}) = \lim_{n\to\infty} \frac{1}{n} H(\varrho_n) = \overline{H}(\hat{\varrho})\,. \tag{7.63}$$

In particular, if the source is memoryless, i.e., $\varrho_n := \varrho^{\otimes n}$ for some quantum state $\varrho \in \mathcal{B}_*(\mathcal{H})$, we have,

$$\underline{H}(\hat{\varrho}) = H(\varrho) = \overline{H}(\hat{\varrho})\,. \tag{7.64}$$

Sketch of Proof: In view of Prop. 7.12, it is enough to show that $\underline{H}(\hat{\varrho}) \geq \overline{H}(\hat{\varrho})$. In the memoryless case, Eq. (7.64) then immediately follows since $H(\varrho^{\otimes n}) = n\,H(\varrho)$ for all $n \in \mathbb{N}$. We assume that $\delta := \frac{1}{2}(\overline{H}(\hat{\varrho}) - \underline{H}(\hat{\varrho})) > 0$. For all $\gamma > \overline{H}(\hat{\varrho})$ we have

$$\overline{\lim_{n\to\infty}}\ \mathrm{tr}\,\{\varrho_n \geq 2^{-n\gamma}\,\mathbb{1}_n\}\varrho_n = 1 \tag{7.65}$$

from the definition of the spectral sup-entropy rate and Prop. 7.6. The quantum asymptotic equipartition theorem [BKS$^+$04, BKS$^+$03] implies that all eigenvalues of ϱ_n are asymptotically concentrated as $n \to \infty$. Hence, for every $\varepsilon > 0$ there exists a positive integer N_ε such that for all $n \geq N_\varepsilon$ we have

$$\mathrm{tr}\,\{\varrho_n \geq 2^{-n(\overline{H}(\hat{\varrho})-\delta)}\,\mathbb{1}_n\}\,\varrho_n \leq \varepsilon\,, \tag{7.66}$$

implying that

$$\overline{\lim_{n\to\infty}}\ \mathrm{tr}\,\{\varrho_n \geq 2^{-n(\overline{H}(\hat{\varrho})-\delta)}\,\mathbb{1}_n\}\,\varrho_n = 0\,. \tag{7.67}$$

However, $\overline{H}(\hat{\varrho}) - \delta = \underline{H}(\hat{\varrho}) + \delta > \underline{H}(\hat{\varrho})$, and hence Eq. (7.67) contradicts the definition of the spectral inf-entropy rate in Eq. (7.51). ■

7.3.4 Mixed Sources

Prop. 7.13 implies that all ergodic quantum information sources are information-stable in the sense of Sec. 7.3.2, and hence a strong converse holds. This Section is devoted to *mixed sources*, which are arguably the simplest non-ergodic quantum sources. We will show below that the strong converse in general fails for these sources.

Given a sequence $(\mathcal{H}_n)_{n\in\mathbb{N}}$ of finite-dimensional Hilbert spaces and two quantum sources $\hat{\sigma} := (\sigma_n)_{n\in\mathbb{N}}$ and $\hat{\omega} := (\omega_n)_{n\in\mathbb{N}}$ with $\sigma_n, \omega_n \in \mathcal{B}_*(\mathcal{H}_n)$, we define the *mixed quantum source* $\hat{\varrho} := (1-p)\,\hat{\sigma} + p\,\hat{\omega}$ by setting

$$\varrho_n := (1-p)\,\sigma_n + p\,\omega_n \tag{7.68}$$

for some *mixing parameter* $p \in [0,1]$. This definition is reminiscent of quantum memory channels with a global classical switch, as introduced in Sec. 6.2.3: with a priori probability p the quantum source $\hat{\omega}$ is chosen, and the output at time n is then given by ω_n.

The source $\hat{\sigma}$ is chosen and applied in every step with the complementary probability $(1-p)$. The information-spectrum techniques allow a straightforward derivation of the compression rate of such a mixed quantum source:

Proposition 7.14. (Data Compression for Mixed Sources)
Let $(\mathcal{H}_n)_{n\in\mathbb{N}}$ be a sequence of finite-dimensional Hilbert spaces. Assume two quantum sources $\hat{\sigma} := (\sigma_n)_{n\in\mathbb{N}}$ and $\hat{\omega} := (\omega_n)_{n\in\mathbb{N}}$ with $\sigma_n, \omega_n \in \mathcal{B}_*(\mathcal{H}_n)$, and let $\hat{\varrho}$ be the corresponding mixed source as defined in Eq. (7.68). We then have,

$$\underline{H}(\hat{\varrho}) = \min\left\{\underline{H}(\hat{\sigma}), \underline{H}(\hat{\omega})\right\} \qquad \text{and} \tag{7.69}$$

$$\overline{H}(\hat{\varrho}) = \max\left\{\overline{H}(\hat{\sigma}), \overline{H}(\hat{\omega})\right\}, \tag{7.70}$$

independently of the mixing parameter $p \in [0,1]$.

For simplicity, in this Section we restrict the discussion to mixed sources with only two component sources. The generalization of Prop. 7.14 to mixed sources with any finite number of components is straightforward. In combination with Th. 7.10, Prop. 7.14 then amounts to a data compression theorem for mixed quantum sources: we may compress the mixed source $\hat{\varrho}$ down to the largest of the individual compression rates. If the component sources are all i.i.d., the direct part of this source coding theorem can also be derived from the universal compression scheme of Jozsa *et al.* [JHH$^+$98], and for ergodic components from the results of Kaltchenko and Yang [KY03]. In contrast, our compression theorem applies more generally to arbitrary component sources, comes with a weak converse, and also allows the formulation of a strong converse: we see from Th. 7.11 that we are guaranteed that the compression errors approach unity only if we try to code at rates smaller than both the individual spectral inf-entropy rates. In contrast to ergodic quantum sources, mixed sources are in general not information-stable: if both sources $\hat{\sigma}$ and $\hat{\omega}$ are i.i.d. with $\sigma_n = \sigma^{\otimes n}$ and $\omega_n = \omega^{\otimes n}$ for two quantum states $\sigma, \omega \in \mathcal{B}_*(\mathcal{H})$, we conclude from Prop. 7.13 and Prop. 7.14 that

$$\underline{H}(\hat{\varrho}) = \min\left\{H(\sigma), H(\omega)\right\} \qquad \text{and} \tag{7.71}$$

$$\overline{H}(\hat{\varrho}) = \max\left\{H(\sigma), H(\omega)\right\}, \tag{7.72}$$

independently of the mixing parameter $p \in [0,1]$. Hence, $\underline{H}(\hat{\varrho}) < \overline{H}(\hat{\varrho})$ whenever $H(\sigma) \neq H(\omega)$.

Proof of Prop. 7.14: We assume without loss that $p \in (0,1)$, for otherwise the statement becomes trivial. Applying Lemma 7.2 and the linearity of the trace, we have

$$\begin{aligned}
&\operatorname{tr}\left\{\varrho_n \geq 2^{-n\gamma}\,\mathbb{1}_n\right\}\left(\varrho_n - 2^{-n\gamma}\,\mathbb{1}_n\right) \\
&\qquad = (1-p)\operatorname{tr}\left(\left\{\varrho_n \geq 2^{-n\gamma}\,\mathbb{1}_n\right\}\left(\sigma_n - 2^{-n\gamma}\,\mathbb{1}_n\right)\right) \\
&\qquad\qquad + p\operatorname{tr}\left(\left\{\varrho_n \geq 2^{-n\gamma}\,\mathbb{1}_n\right\}\left(\omega_n - 2^{-n\gamma}\,\mathbb{1}_n\right)\right) \\
&\qquad \leq (1-p)\operatorname{tr}\left(\left\{\sigma_n \geq 2^{-n\gamma}\,\mathbb{1}_n\right\}\left(\sigma_n - 2^{-n\gamma}\,\mathbb{1}_n\right)\right) \\
&\qquad\qquad + p\operatorname{tr}\left(\left\{\omega_n \geq 2^{-n\gamma}\,\mathbb{1}_n\right\}\left(\omega_n - 2^{-n\gamma}\,\mathbb{1}_n\right)\right)
\end{aligned} \tag{7.73}$$

for any $\gamma \in \mathbb{R}$. Setting now $\gamma := \overline{H}(\hat{\varrho}) + \delta$ for some $\delta > 0$ and taking the $\overline{\lim}$, the LHS of Eq. (7.73) approaches 1, and hence the same holds true for both traces on the RHS. Since $\delta > 0$ is arbitrary, we conclude from the definition of the spectral sup-entropy rate that

$$\overline{H}(\hat{\varrho}) \geq \max\left\{\overline{H}(\sigma), \overline{H}(\omega)\right\}. \tag{7.74}$$

If instead we set $\gamma := \min\{\underline{H}(\sigma), \underline{H}(\omega)\} - \delta$ and again take the $\overline{\lim}$, the RHS of Eq. (7.73) vanishes, implying that

$$\underline{H}(\hat{\varrho}) \geq \min\left\{\underline{H}(\sigma), \underline{H}(\omega)\right\}. \tag{7.75}$$

In order to prove the converse implications, we explicitly construct a sequence of projection operators: For $\beta \in \mathbb{R}$ we set $P_n^0 := \{\sigma_n \geq 2^{-n\beta}\, \mathbb{1}_n\}$ and $Q_n := \{\omega_n \geq 2^{-n\beta}\, \mathbb{1}_n\}$. We assume that Q_n has the spectral decomposition $Q_n = \sum_{i=1}^{q} |i\rangle\langle i|$ with $q := \operatorname{tr} Q_n$. Starting with P_n^0, we now iteratively define a sequence of projection operators $\{P_n^i\}_{i=0}^{q}$ as follows: For each $i = 1, ..., q$, if $|i\rangle$ lies entirely in the subspace onto which P_n^{i-1} projects, we set $P_n^i := P_n^{i-1}$. Otherwise, we take the component of $|i\rangle$ orthogonal to that subspace, say $|i^\perp\rangle$, and we let $P_n^i := P_n^{i-1} \oplus |i^\perp\rangle\langle i^\perp|$. From Lemma 7.3 we conclude that

$$\max\left\{\operatorname{tr}\left\{\sigma_n \geq 2^{-n\beta}\, \mathbb{1}_n\right\}, \operatorname{tr}\left\{\sigma_n \geq 2^{-n\beta}\, \mathbb{1}_n\right\}\right\} \leq 2^{n\beta}, \tag{7.76}$$

implying that $\operatorname{tr} P_n^q \leq 2 \cdot 2^{n\beta}$. We can now apply Lemma 7.2 to obtain for all $\gamma \in \mathbb{R}$,

$$\begin{aligned}
&\operatorname{tr}\left\{\varrho_n \geq 2^{-n\gamma}\, \mathbb{1}_n\right\}(\varrho_n - 2^{-n\gamma}\, \mathbb{1}_n) \\
&\quad \geq \operatorname{tr} P_n^q\, (\varrho_n - 2^{-n\gamma}\, \mathbb{1}_n) \\
&\quad = (1-p) \operatorname{tr} P_n^q\, (\sigma_n - 2^{-n\gamma}\, \mathbb{1}_n) + p \operatorname{tr} P_n^q\, (\omega_n - 2^{-n\gamma}\, \mathbb{1}_n) \\
&\quad \geq (1-p) \operatorname{tr}\left\{\sigma_n \geq 2^{-n\gamma}\, \mathbb{1}_n\right\}\sigma_n + p \operatorname{tr}\left\{\omega_n \geq 2^{-n\gamma}\, \mathbb{1}_n\right\}\omega_n - 2^{-n\gamma} \operatorname{tr} P_n^q \\
&\quad \geq (1-p) \operatorname{tr}\left\{\sigma_n \geq 2^{-n\gamma}\, \mathbb{1}_n\right\}\sigma_n + p \operatorname{tr}\left\{\omega_n \geq 2^{-n\gamma}\, \mathbb{1}_n\right\}\omega_n - 2 \cdot 2^{-n(\gamma-\beta)} \\
&\quad \geq (1-p) \operatorname{tr}\left\{\sigma_n \geq 2^{-n\gamma}\, \mathbb{1}_n\right\}(\sigma_n - 2^{-n\gamma}\, \mathbb{1}_n) \\
&\qquad\quad + p \operatorname{tr}\left\{\omega_n \geq 2^{-n\gamma}\, \mathbb{1}_n\right\}(\omega_n - 2^{-n\gamma}\, \mathbb{1}_n) - 2 \cdot 2^{-n(\gamma-\beta)}.
\end{aligned} \tag{7.77}$$

Setting $\beta := \max\{\overline{H}(\hat{\sigma}), \overline{H}(\hat{\omega})\} + \delta$ for some $\delta > 0$ and $\gamma := \beta + \delta$ and then taking the $\underline{\lim}$, the RHS of Eq. (7.77) approaches 1. Since δ is arbitrary, we may thus conclude from the definition of the spectral sup-entropy rate that

$$\overline{H}(\hat{\varrho}) \leq \max\left\{\overline{H}(\hat{\sigma}), \overline{H}(\hat{\omega})\right\}. \tag{7.78}$$

If instead we choose $\gamma := \underline{H}(\hat{\varrho}) - \delta$ and $\beta := \gamma - \delta = \underline{H}(\hat{\varrho}) - 2\,\delta$ for some $\delta > 0$ and again take the $\overline{\lim}$, the LHS of Eq. (7.77) vanishes by the definition of the spectral inf-entropy rate. Hence, both traces on the RHS of Eq. (7.77) must likewise vanish in this limit, implying that

$$\underline{H}(\hat{\varrho}) \leq \min\left\{\underline{H}(\hat{\sigma}), \underline{H}(\hat{\omega})\right\}, \tag{7.79}$$

since $\delta > 0$ was arbitrary. Eq. (7.69) now follows from combining Eqs. (7.75) and (7.79), and Eq. (7.70) follows from Eqs. (7.74) and (7.78). ■

7.4 Summary and Outlook

In this Chapter we have presented quantum information spectrum methods as an alternative approach to coding theorems for general (non-i.i.d.) sources and channels. The focus of our exposition has mostly been on source coding: we have proven a data compression theorem for general quantum sources (without any structural assumptions), in which the compression rate is expressed in terms of the spectral sup-entropy rate. We have then applied this theorem to ergodic as well as mixed (non-ergodic) sources.

The spectral divergence rates (of which the spectral entropy rates are a special case) can be seen as a generalization of the quantum relative entropy and make their appearance in all known coding theorems. Hayashi has shown [Hay03, Hay06^a] that the spectral inf-entropy rate describes the optimal asymptotic rate at which *pure* state can be converted into maximally entangled states by means of local operations and classical communication (LOCC): the distillable entanglement of a sequence of pure bipartite states $(\psi_n)_{n\in\mathbb{N}}$ with $|\psi_n\rangle \in \mathcal{H}_n^A \otimes \mathcal{H}_n^B$ equals $\underline{H}(\hat{\varrho})$, where the source $\hat{\varrho} \equiv (\varrho_n)_{n\in\mathbb{N}}$ is determined by the local restriction, $\varrho_n := \mathrm{tr}_B\, |\psi_n\rangle\langle\psi_n|$. We have seen in Prop. 7.13 that for $\psi_n = \psi^{\otimes n}$ the inf-entropy rate indeed just coincides with the von Neumann entropy, the familiar result for the distillable entanglement of a bipartite pure state $|\psi\rangle\langle\psi| \in \mathcal{B}_*(\mathcal{H}^A \otimes \mathcal{H}^B)$.

Hayashi and Nagaoka [HN03] have also derived an information-spectrum characterization for the classical capacity of a sequence of general quantum channels: Let $(\mathcal{H}_n)_{n\in\mathbb{N}}$ and $(\mathcal{K}_n)_{n\in\mathbb{N}}$ be two sequences of finite-dimensional Hilbert spaces, and let $\hat{T} := (T_n)_{n\in\mathbb{N}}$ be a sequence of quantum channels such that $T_{n*}\colon \mathcal{B}_*(\mathcal{H}_n) \to \mathcal{B}_*(\mathcal{K}_n)$. Again we do not make any structural assumptions on the interrelation of the individual channels T_n. The capacity $C(\hat{T})$ of the sequence $\hat{T}$ for classical information transmission can be defined in complete analogy to the memoryless case described in Sec. 5.4.1, replacing only the n-fold tensor product $T^{\otimes n}$ by T_n in every step. Hayashi and Nagaoka were able to show that $C(\hat{T})$ can be expressed in terms of the spectral inf-mutual information rate, $\underline{H}(\hat{\varrho}^A : \hat{\varrho}^B) := -\overline{D}(\hat{\varrho} \parallel \hat{\varrho}^A \otimes \hat{\varrho}^B)$, as follows:

$$C(\hat{T}) = \sup_{\hat{\varrho}\in\mathcal{S}_{cq}} \underline{H}\big(\hat{\varrho}^A : \hat{T}(\hat{\varrho}^B)\big)\,. \tag{7.80}$$

The maximization in Eq. (7.80) is performed over all classical-quantum input states of the form

$$\mathcal{S}_{cq} \ni \varrho_n := \sum_i p_i\, |i\rangle\langle i|^A \otimes \varrho_{n,i}^B\,, \tag{7.81}$$

where $\{p_i\}_i$ is a classical probability distribution, and $\varrho_{n,i} \in \mathcal{B}_*(\mathcal{H}_n)$ are quantum states on the input space $\mathcal{H}_n$. Again, for memoryless channels this reduces to the well-known Holevo-Schumacher-Westmoreland theorem [Hol98, SW97], Th. 5.5 in Sec. 5.5.1.

The quantum channel capacity $Q(\hat{T})$ of the sequence $\hat{T} := (T_n)_{n\in\mathbb{N}}$ can likewise be defined along the lines of Def. 5.1. Extrapolating from the i.i.d. case, one would expect

that $Q(\hat{T})$ can be expressed in terms of the spectral sup-conditional information rate

$$\overline{H}(\hat{\varrho}^A|\hat{T}(\hat{\varrho}^B)) := -\underline{D}(\hat{\varrho} \parallel \mathbb{1}^A \otimes \hat{T}(\hat{\varrho}^B)) \,. \tag{7.82}$$

However, so far this can be demonstrated only for special cases, which all involve some degree of commutativity, and a general coding theorem for the quantum channel capacity remains open.

Even when general coding theorems for states and channels in terms of spectral quantities can be derived, their conversion into more accessible information-theoretic quantities is often a mayor challenge. Even in the completely classical setting, to date the only non-ergodic examples for which this has been done convincingly seem to be mixed sources and channels [Han03]. So even in the purely classical setting there remain many challenges for future research in order for the information spectrum techniques to emerge as a truly powerful toolkit in information theory.

Chapter 8

Continuity of Distillable Entanglement

Distillable entanglement quantifies the optimal asymptotic rate at which many copies of a given bipartite quantum state can be converted into maximally entangled qubit pairs by local quantum operations and classical communication. Together with the quantum and classical channel capacities discussed in Ch. 5, distillable entanglement lies at the heart of quantum information theory. In this Chapter we investigate the continuity of distillable entanglement, and we show uniform norm bounds for all states with full support. We conjecture that uniform continuity also holds at the boundary of the state space, and hence everywhere, though we have not been able to demonstrate this yet. Our results also imply the continuity of the quantum channel capacity assisted by 2-way classical side channels on the boundary of the channels with vanishing capacity.

The results described in this Chapter are joint work with Matthias Christandl.

8.1 Introduction and Overview

We have seen in Sec. 1.1 that entanglement is a precious physical resource. It allows to generate secure cryptographic keys [Eke91], helps to teleport quantum states over classical channels [BBC$^+$93], and may double the capacity of classical information channels [BW92]. Most of such protocols — in particular long distance quantum key distribution involving a quantum repeater — require entanglement in the form of maximally entangled qubit pairs, or *ebits*. But due to decoherence or technological constraints such ebits are usually not readily available. We may then try to distill them from a given (possibly mixed and usually less entangled) bipartite quantum state ϱ that we know how to prepare in the lab. The *distillable entanglement* characterizes how well this transformation may be performed in the limit of asymptotically many copies $\varrho^{\otimes n}$, using only local quantum operations and classical communication (LOCC) in the process. It

is one of the key quantitative notions of quantum information theory.

Definition 8.1. (Distillable Entanglement)
Let $\mathcal{H}_A$ and $\mathcal{H}_B$ be finite-dimensional Hilbert spaces, and $k \in \{1, 2\}$. A positive number r is called a k-way achievable distillation rate *for the bipartite state $\varrho \in \mathcal{B}_*(\mathcal{H}_A \otimes \mathcal{H}_B)$ iff there exists an integer sequence $(n_\nu)_{\nu\in\mathbb{N}}$ with $\lim_{\nu\to\infty} n_\nu = \infty$ and a k-way LOCC protocol $(\Lambda^k_{n_\nu})_{\nu\in\mathbb{N}}$ such that*

$$\lim_{\nu\to\infty} \|\Lambda^k_{n_\nu}(\varrho^{\otimes n_\nu}) - |\Omega\rangle\langle\Omega|^{\otimes r n_\nu}\|_1 = 0\,, \tag{8.1}$$

where $|\Omega\rangle \equiv |\Omega_2\rangle := 1/\sqrt{2}(|00\rangle + |11\rangle)$ is a maximally entangled qubit pair. The index $k \in \{1, 2\}$ in Eq. (8.1) denotes whether the supremum is taken over 1-way (from Alice to Bob) or 2-way LOCC operations. The distillable entanglement *$D_k(\varrho)$ is defined as the supremum of the achievable distillation rates.*

Remark 8.2. Since we can always embed the smaller of the Hilbert spaces $\mathcal{H}_A$ and $\mathcal{H}_B$ into the larger one, for simplicity we will henceforth assume that $\mathcal{H}_A = \mathcal{H}_B =: \mathcal{H}$, and we will set $d := \dim \mathcal{H}$.

Remark 8.3. The distillable entanglement is defined very much in analogy to the quantum channel capacity in Def. 5.1 and the optimal source compression rate in Def. 7.7, and several of the remarks we made then are seen to apply to the distillable entanglement as well. In particular, in lieu of the "optimistic" Def. 8.1 we could have required that the trace norm distance in Eq. (8.1) vanishes asymptotically along *any* sequence $(n_\nu)_{\nu\in\mathbb{N}}$ that achieves the rate r. It follows from the results of Devetak and Winter [DW04[a]] that both concepts again lead to equivalent definitions, but we will not go into the details here and instead refer to Hayashi's text [Hay06[b]].

Remark 8.4. Obviously, $D_2(\varrho) \geq D_1(\varrho)$. There are states for which this inequality is strict: the two-qubit Werner state $\varrho = 1/2|\Omega\rangle\langle\Omega| + 1/8\,\mathbb{1}$ is 2-way distillable, but not 1-way distillable [BDS$^+$96].

If $\varrho \equiv |\psi\rangle\langle\psi|$ is pure, entanglement distillation is sometimes called *entanglement concentration*. In this case, 1-way and 2-way distillable entanglement coincide, and simply equal the von Neumann entropy of the local restriction [BBP$^+$96[b]],

$$D_1(|\psi\rangle\langle\psi|) = H(\varrho_A) = H(\varrho_B) = D_2(|\psi\rangle\langle\psi|)\,, \tag{8.2}$$

where $\varrho_B := \mathrm{tr}_A\,\varrho$ denotes the restriction of the bipartite state $\varrho \in \mathcal{B}_*(\mathcal{H}_A) \otimes \mathcal{B}_*(\mathcal{H}_B)$ to Bob's subsystem B, and analogously for the restriction $\varrho_A := \mathrm{tr}_B\,\varrho \in \mathcal{B}_*(\mathcal{H}_A)$.

For the general case, Devetak and Winter [DW04[a]] have shown that both D_1 and D_2 may be expressed in terms of the *coherent information* $I_c(\varrho) = H(\varrho_B) - H(\varrho)$ (cf. Appendix A),

$$D_k(\varrho) = \lim_{n\to\infty} \frac{1}{n} \sup_{\Lambda^k_n} I_c\big(\Lambda^k_n(\varrho^{\otimes n})\big)\,. \tag{8.3}$$

For the 2-way distillable entanglement $D_2(\varrho)$, the supremum in Eq. (8.3) is over all 2-way LOCC operations Λ_n^2. For the 1-way distillable entanglement, the supremum is restricted to the quantum instruments Λ_n^1 on Alice's side (cf. Sec. 2.4).

Just as for quantum channel capacities, calculating the distillable entanglement straight from Def. 8.1 or even the simplified entropic expressions Eq. (8.3) is a tricky business, since a limit over LOCC operations on asymptotically large quantum systems is involved. For the same reason, determining global properties of D_1 and D_2 has often proved a difficult task. Both additivity and convexity fail to hold for D_2 provided that bound entangled Werner states with negative partial transpose (NPT) exist [SST01] — a long-standing open conjecture.

In this Chapter we will show that even though they may not be convex, both the 1-way and 2-way distillable entanglement are uniformly continuous on the set of density operators in trace norm topology — at least on the (dense) set of states with full support. Hence, if two quantum states are almost indistinguishable the amount of entanglement that can be asymptotically distilled from these states is nearly the same. Since any state preparation in the lab necessarily involves some finite errors, such a continuity property is crucial for an unambiguous operational interpretation of the distillable entanglement. Our results also imply that the quantum channel capacity assisted by 2-way classical side channels is continuous in the neighborhood of all channels with vanishing capacity.

Preliminary results on the continuity of distillable entanglement, with weaker bounds and several flaws at crucial points, can be found in Vidal's preprint [Vid02].

As a warm-up exercise, we show in Sec. 8.2 that both the 1-way and the 2-way distillable entanglement are uniformly continuous in the neighborhood of all pure quantum states. In Sec. 8.3 we will then make use of the entropic expressions Eq. (8.3) to give a simple proof of the lower semi-continuity of both the 1-way and the 2-way distillable entanglement. An estimate in the converse direction will be presented in Sec. 8.4 — albeit for convex mixtures only. Our methods require the 2-way distillable entanglement to be non-zero. Hence, the boundary of the set of 2-way distillable states demands a separate treatment, which will be performed in Sec. 8.5. If the quantum state under consideration is *faithful*, i.e., has full support, continuity on convex mixtures is sufficient to show continuity for all states, as explained in Sec. 8.6. In Sec. 8.7 we apply our continuity results to show that the quantum channel capacity assisted by 2-way classical side channels is likewise uniformly continuous on the boundary of channels with vanishing capacity. We conclude with a Summary and Outlook in Sec. 8.8.

8.2 Pure States

We will show in this Section that both the 1-way and the 2-way distillable entanglement are continuous in the neighborhood of the pure states. From the *hashing inequality*

[HHH00, DW04[a]], we know that $D_1(\varrho) \geq I_c(\varrho)$ for all quantum states $\varrho \in \mathcal{B}_*(\mathcal{H}_A \otimes \mathcal{H}_B)$. Moreover, the 2-way distillable entanglement $D_2(\varrho)$ is clearly upper bounded by the entanglement cost $E_C(\varrho)$, and hence by the *entanglement of formation* [BDS+96, PV07],

$$E_F(\varrho) := \inf \Big\{ \sum_i p_i \, H(\varrho_i^A) \;\Big|\; \varrho = \sum_i p_i \, |\psi_i\rangle\langle\psi_i| \Big\} \,. \tag{8.4}$$

The infimum in Eq. (8.4) is taken over all decompositions of $\varrho \in \mathcal{B}_*(\mathcal{H}_A \otimes \mathcal{H}_B)$ into pure states $|\psi_i\rangle\langle\psi_i|$, with $\varrho_i^A := \mathrm{tr}_B |\psi_i\rangle\langle\psi_i|$. We then conclude from the concavity of the entropy (cf. Prop. A.4) that $E_F(\varrho) \leq H(\varrho^A)$ with $\varrho^A := \mathrm{tr}_B \varrho$. In summary, we have the following chain of inequalities,

$$I_c(\varrho) \leq D_1(\varrho) \leq D_2(\varrho) \leq E_C(\varrho) \leq E_F(\varrho) \leq H(\varrho^A) \,. \tag{8.5}$$

Now if $\varrho \equiv |\psi\rangle\langle\psi|$ is pure, we know from our discussion in Sec. 8.1 that all the quantities in Eq. (8.5) coincide. Given a quantum state $\sigma \in \mathcal{B}_*(\mathcal{H}_A \otimes \mathcal{H}_B)$ with $\| |\psi\rangle\langle\psi| - \sigma \|_1 \leq \varepsilon$, we may then apply the continuity bounds from Prop. A.6 and Prop. A.7 to conclude that

$$\begin{aligned} D_2(\sigma) \geq D_1(\sigma) \geq I_c(\sigma) &\geq I_c(|\psi\rangle\langle\psi|) - f_1(\varepsilon) \\ &= D_1(|\psi\rangle\langle\psi|) - f_1(\varepsilon) \\ &= D_2(|\psi\rangle\langle\psi|) - f_1(\varepsilon) \\ &= H(\mathrm{tr}_B |\psi\rangle\langle\psi|) - f_1(\varepsilon) \\ &\geq H(\sigma^A) - f_1(\varepsilon) - f_2(\varepsilon) \\ &\geq D_2(\sigma) - f_1(\varepsilon) - f_2(\varepsilon) \\ &\geq D_1(\sigma) - f_1(\varepsilon) - f_2(\varepsilon) \,, \end{aligned} \tag{8.6}$$

where for notational convenience we have introduced the shorthands

$$f_1(s) := 4\, s \lg d + 2\, \eta(s) + 2\, \eta(1-s) \qquad \text{and} \tag{8.7}$$

$$f_2(s) := 2\, s \lg d + \eta(s) \tag{8.8}$$

for $\eta(s) := -s \lg s$, and $d := \dim \mathcal{H}_A = \dim \mathcal{H}_B$. Since both functions $f_1(s)$ and $f_2(s)$ vanish continuously as $s \to 0$, continuity of both the 1-way and the 2-way distillable entanglement immediately follow from Eq. (8.6). We summarize this result as

Proposition 8.5. (Continuity near Pure States)
Let $|\psi\rangle \in \mathcal{H} \otimes \mathcal{H}$. We then have for all states $\sigma \in \mathcal{B}_*(\mathcal{H} \otimes \mathcal{H})$ with $\| |\psi\rangle\langle\psi| - \sigma \|_1 \leq \varepsilon$,

$$|D_k(|\psi\rangle\langle\psi|) - D_k(\sigma)| \leq 6\, \varepsilon \lg d + 3\, \eta(\varepsilon) + 2\, \eta(1-\varepsilon) \,, \tag{8.9}$$

where $\eta(s) := -s \lg s$ and $d := \dim \mathcal{H}$.

8.3 Lower Semi-Continuity

In this Section we will show that both the 1-way and the 2-way distillable entanglement are lower semi-continuous. As a reminder, we start with a definition [RS80]:

Definition 8.6. (Lower Semi-Continuity)
Let X be a topological space. A function $f: X \to \mathbb{R}$ is called lower semi-continuous *at $x_0 \in X$ iff for all $\varepsilon > 0$, there is a neighborhood U of x_0 such that $f(x) \geq f(x_0) - \varepsilon$ for all $x \in U$. f is said to be* lower semi-continuous *iff it is lower semi-continuous at all points in its domain.*

When applied to the function $f = D_k$, this definition means that for every quantum state $\varrho \in \mathcal{B}_*(\mathcal{H} \otimes \mathcal{H})$ and $\varepsilon > 0$ we may find a positive constant $\delta \equiv \delta(\varrho, \varepsilon)$ such that $D_k(\sigma) \geq D_k(\varrho) - \varepsilon$ whenever $\|\varrho - \sigma\|_1 \leq \delta$.

Proposition 8.7. *Both D_1 and D_2 are lower semi-continuous.*

Proof: It is clear from the entropic expressions in Eq. (8.3) that we may find a positive integer $N \equiv N(\varrho, \varepsilon, k) \in \mathbb{N}$ and a corresponding LOCC protocol $(\Lambda_n^k)_{n \in \mathbb{N}}$ such that

$$D_k(\varrho) \geq \frac{1}{n} I_c\big(\Lambda_n^k(\varrho^{\otimes n})\big) \geq D_k(\varrho) - \varepsilon \tag{8.10}$$

for all $n \geq N$. For all quantum states $\sigma \in \mathcal{B}_*(\mathcal{H} \otimes \mathcal{H})$, $\|\varrho - \sigma\|_1 \leq \frac{\varepsilon}{n}$ implies that $\|\varrho^{\otimes n} - \sigma^{\otimes n}\|_1 \leq \varepsilon$, and hence $\|\Lambda_n^k(\varrho^{\otimes n}) - \Lambda_n^k(\sigma^{\otimes n})\|_1 \leq \varepsilon$ from the monotonicity of the trace norm under quantum operations. We may then conclude from Prop. A.7 that

$$\begin{aligned} D_k(\varrho) - D_k(\sigma) &\leq \varepsilon + \frac{1}{n}\Big(I_c\big(\Lambda_n^k(\varrho^{\otimes n})\big) - I_c\big(\Lambda_n^k(\sigma^{\otimes n})\big)\Big) \\ &\leq \varepsilon + \frac{1}{n}\big(4\,\|\Lambda_n^k(\varrho^n) - \Lambda_n^k(\sigma^n)\|_1\,\operatorname{ld} d^n + 2\big) \\ &\leq \varepsilon + 4\,\varepsilon \operatorname{ld} d + \frac{2}{n} \end{aligned} \tag{8.11}$$

for all $n \geq N$, where again we have set $d := \dim \mathcal{H}$. Choosing n sufficiently large such that $\frac{2}{n} \leq \varepsilon$, we see from Eq. (8.11) that

$$D_k(\sigma) \geq D_k(\varrho) - 2\,\varepsilon - 4\,\varepsilon\,\operatorname{ld} d \tag{8.12}$$

for all quantum states $\sigma \in \mathcal{B}_*(\mathcal{H} \otimes \mathcal{H})$ satisfying $\|\varrho - \sigma\|_1 \leq \frac{\varepsilon}{n}$ for some sufficiently large n. Hence, both D_1 and D_2 are lower semi-continuous. ■

Prop. 8.7 shows that for every state with non-vanishing distillable entanglement we may find a finite-size neighborhood in which all states are likewise distillable. In other words, the set of distillable states,

$$\mathcal{D}_k := \{\sigma \in \mathcal{B}_*(\mathcal{H} \otimes \mathcal{H}) \;\mid\; D_k(\sigma) > 0\}\,, \tag{8.13}$$

is open. We summarize this result as a corollary:

Corollary 8.8. *The sets $\mathcal{D}_k$ are open in trace-norm topology.*

8.4 A Lower Bound for Convex Mixtures

In Sec. 8.3 we have derived an upper bound on the distillable entanglement $D_k(\varrho)$ of a quantum state $\varrho \in \mathcal{B}_*(\mathcal{H}\otimes\mathcal{H})$ in terms of the distillable entanglement of the states in its neighborhood. Here we prove a corresponding lower bound. By a result of Alicki and Fannes [AF04], it is sufficient to consider convex mixtures of the form $\gamma = (1-\varepsilon)\varrho+\varepsilon\sigma$:

Lemma 8.9. (Alicki & Fannes)
Let $\varrho, \tau \in \mathcal{B}_*(\mathcal{H} \otimes \mathcal{H})$ be two quantum states with trace norm difference $\varepsilon := \|\varrho - \tau\|_1$. Then we may find two quantum states $\sigma_1, \sigma_2 \in \mathcal{B}_*(\mathcal{H} \otimes \mathcal{H})$ such that

$$(1-\varepsilon)\,\varrho + \varepsilon\,\sigma_1 = (1-\varepsilon)\,\tau + \varepsilon\,\sigma_2\,. \tag{8.14}$$

How are we going to use this Lemma? Assume that we have established continuity of distillable entanglement on mixtures, i. e.,

$$|D_k(\varrho) - D_k(\gamma)| \leq f_k(\varepsilon) \tag{8.15}$$

for $\gamma = (1-\varepsilon)\varrho+\varepsilon\sigma$ and some function f_k such that $\lim_{\varepsilon\to 0} f_k(\varepsilon) = 0$. Then Lemma 8.9 in conjunction with the triangle inequality will allow to extent the continuity to the general case: for any two quantum states ϱ, τ with $\|\varrho - \tau\|_1 \leq \varepsilon$, we obtain

$$|D_k(\varrho) - D_k(\tau)| \leq |D_k(\varrho) - D_k(\gamma)| + |D_k(\gamma) - D_k(\tau)| \leq 2\, f_k(\varepsilon)\,. \tag{8.16}$$

Thus, we will henceforth focus on convex mixtures only.

Proposition 8.10. *Let $\rho \in \mathcal{B}_*(\mathcal{H} \otimes \mathcal{H})$ and $\gamma := (1-\varepsilon)\varrho + \varepsilon\sigma$ for some $\varepsilon > 0$ and $\sigma \in \mathcal{B}_*(\mathcal{H} \otimes \mathcal{H})$. Then*

$$D_1(\gamma) \leq (1-\varepsilon)\, D_1(\rho) + \varepsilon \operatorname{ld} d\,, \tag{8.17}$$

where $d := \dim \mathcal{H}$. The same bound holds for D_2, provided $D_2(\varrho) > 0$.

The proof of Prop. 8.10 relies in part on Vidal's work [Vid02] on the asymptotic mixing of quantum states, which we cite here without proof:

Lemma 8.11. (Asymptotic Mixing)
Let $\rho, \sigma \in \mathcal{B}_*(\mathcal{H} \otimes \mathcal{H})$. For every $p \in [0,1]$ and $\delta > 0$ there is a positive integer $N \equiv N(\delta, p)$ and a 1-way LOCC protocol $(\Lambda_n^1)_{n\in\mathbb{N}}$ such that

$$\|\Lambda_n^1\big(\varrho^{\otimes(1-p)n+o(n)} \otimes \sigma^{\otimes pn+o(n)}\big) - \big((1-p)\varrho + p\sigma\big)^{\otimes n}\|_1 \leq \delta \qquad \forall \quad n \geq N. \tag{8.18}$$

The coherent information is convex [LR73], yet not too much:

Lemma 8.12. *For a finite collection of bipartite quantum states $\varrho_i \in \mathcal{B}_*(\mathcal{H}_A \otimes \mathcal{H}_B)$ with a corresponding probability distribution $\{p_i\}_i$, we have*

$$\sum_i p_i\, I_c(\varrho_i) - I_c(\varrho) \leq H(\varrho) - \sum_i p_i\, H(\varrho_i) \leq H(\{p_i\}_i)\,, \tag{8.19}$$

where $\varrho := \sum_i p_i\, \varrho_i$ denotes the averaged state, and $H(\{p_i\}_i)$ is the Shannon entropy of the classical distribution $\{p_i\}_i$.

Proof: The proof of Lemma 8.12 is immediate from Prop. A.4,

$$\begin{aligned}\sum_i p_i\, I_c(\varrho_i) - I_c(\varrho) &= \sum_i p_i\, H(\varrho_i^B) - \sum_i p_i\, H(\varrho_i) - H(\varrho^B) + H(\varrho) \\ &\leq H(\varrho) - \sum_i p_i\, H(\varrho_i) \leq H(\{p_i\}_i)\,. \ \blacksquare\end{aligned} \tag{8.20}$$

We can apply this Lemma to show that the 1-way distillable entanglement is additive under tensoring with maximally entangled states:

Lemma 8.13. *For any quantum state $\varrho \in \mathcal{B}_*(\mathcal{H} \otimes \mathcal{H})$ we have*

$$D_1(\varrho \otimes |\Omega\rangle\langle\Omega|) = D_1(\varrho) + 1\,. \tag{8.21}$$

Proof: Obviously, $D_1(\varrho \otimes |\Omega\rangle\langle\Omega|) \geq D_1(\varrho) + 1$. In order to show the converse implication, we define the quantum states

$$\alpha_{i_1\ldots i_n} := \rho^{\otimes n} \otimes (\mathbb{1}^A \otimes \hat{\sigma}^B_{i_1\ldots i_n})|\Omega\rangle\langle\Omega|^{\otimes n}(\mathbb{1}^A \otimes \hat{\sigma}^B_{i_1\ldots i_n})\,, \tag{8.22}$$

where $i_j \in \{0,1\}$ for all $j = 1,\ldots,n$, and $\hat{\sigma}^B_{i_1\ldots i_n} := \sigma_{i_1} \otimes \cdots \otimes \sigma_{i_n}$ is an n-fold tensor product of $\sigma_0 = \mathrm{id}$ and the Pauli matrix $\sigma_1 = \sigma_z$. Sampling uniformly at random from all the states $\alpha_{i_1\ldots i_n}$ implements the dephasing channel channel [HHH$^+$05], and hence

$$2^{-n} \sum_{i_1\ldots i_n} \alpha_{i_1\ldots i_n} = \rho^{\otimes n} \otimes \tau^{\otimes n}\,, \tag{8.23}$$

where $\tau = \frac{|00\rangle\langle 00| + |11\rangle\langle 11|}{2}$.

Now let $(\Lambda^1_n)_{n\in\mathbb{N}}$ be a sequence of quantum instruments, as in Eq. (8.3). They commute with the local unitary transformations $\hat{\sigma}^B_{i_1\ldots i_n}$ on Bob's side. (This will in general fail to hold for 2-way LOCC operations.) Since the coherent information is invariant under such unitary transformations, we can conclude from Lemma 8.12 that

$$\begin{aligned} I_c\big(\Lambda^1_n(\rho^{\otimes n} \otimes |\Omega\rangle\langle\Omega|^{\otimes n})\big) &= I_c\big(\Lambda^1_n(\alpha_{0\ldots 0})\big) \\ &= 2^{-n} \sum_{i_1\ldots i_n} I_c\big(\Lambda^1_n(\alpha_{i_1\ldots i_n})\big) \\ &\leq I_c\big(\Lambda^1_n(\rho^{\otimes n} \otimes \tau^{\otimes n})\big) + n\,. \end{aligned} \tag{8.24}$$

Choosing Λ^1_n as the instrument that attains the supremum in Eq. (8.3) for the quantum state $\varrho \otimes |\Omega\rangle\langle\Omega|$, dividing by n and letting $n \to \infty$, we infer from Eq. (8.24) that

$$D_1(\varrho \otimes |\Omega\rangle\langle\Omega|) \leq D_1(\varrho \otimes \tau) + 1 = D_1(\varrho) + 1\,. \tag{8.25}$$

The last equality holds because τ is separable. $\blacksquare$

The proof of Lemma 8.13 shows that a sublinear amount of pre-shared ebits cannot help Alice and Bob to increase the 1-way distillable entanglement $D_1(\varrho)$. We summarize this result as a Corollary:

Corollary 8.14. (Entanglement-Assisted Distillable Entanglement) *For a quantum state $\rho \in \mathcal{B}_*(\mathcal{H}\otimes\mathcal{H})$, let $D_1^{sub}(\varrho)$ be defined exactly as the 1-way distillable entanglement in Def. 8.1 above, but with a sublinear amount of pre-shared ebits as an additional resource. We then have*

$$D_1^{sub}(\varrho) = D_1(\varrho)\,. \tag{8.26}$$

We now introduce the auxiliary states

$$\gamma' := (1-\varepsilon)|0\rangle\langle 0| \otimes \rho + \varepsilon|1\rangle\langle 1| \otimes \sigma \qquad \text{and} \tag{8.27}$$

$$\gamma'' := (1-\varepsilon)|0\rangle\langle 0| \otimes \rho + \varepsilon|1\rangle\langle 1| \otimes |\Omega_d\rangle\langle\Omega_d|\,, \tag{8.28}$$

where $|\Omega_d\rangle := \frac{1}{\sqrt{d}}\sum_{i=1}^d |ii\rangle$ is maximally entangled, and $\{|i\rangle\langle i|\}_{i=0,1}$ are orthogonal basis states of a qubit flag system on Alice's side. The introduction of this flag system facilitates the calculation of the distillable entanglement:

Lemma 8.15. *For γ'' defined as in Eq. (8.28) we have*

$$D_1(\gamma'') = (1-\varepsilon)D_1(\rho) + \varepsilon \operatorname{ld} d\,. \tag{8.29}$$

If $D_2(\varrho) > 0$, the same equation holds for D_2.

Proof: We first show that $D_k(\gamma'') \geq (1-\varepsilon)D_k(\rho) + \varepsilon \operatorname{ld} d$ for $k \in \{1,2\}$. Starting with an n-fold tensor product $\gamma''^{\otimes n}$, Alice performs local measurements on the flag system, and communicates the measurement outcome to Bob. The flag system is then discarded. We denote this 1-way LOCC operation by Λ_n^1, so the resulting output state is $\Lambda_n^1(\gamma''^{\otimes n})$. The classical law of large numbers guarantees that for any $\delta > 0$ there exists a positive integer $N \equiv N(\delta)$ such that

$$\|\Lambda_n^1(\gamma''^{\otimes n}) - \varrho^{\otimes\lfloor(1-\varepsilon)n\rfloor} \otimes |\Omega_d\rangle\langle\Omega_d|^{\otimes\lfloor\varepsilon n\rfloor}\|_1 \leq \delta \qquad \forall \;\; \geq N\,. \tag{8.30}$$

Let $(\Xi_n^k)_{n\in\mathbb{N}}$ be a k-way distillation protocol for the state ϱ with distillation rate r, and Γ_n^d the local operation that reorders qudits into qubits. Application of $\Xi_n^k \otimes \Gamma_n^d$ does not increase the trace norm, and thus we have

$$\|\big(\Xi_n^k \otimes \Gamma_n^d\big) \circ \Lambda_n^1(\gamma''^{\otimes n}) - |\Omega\rangle\langle\Omega|^{\otimes\lfloor n(1-\varepsilon)r + n\varepsilon \operatorname{ld} d\rfloor}\|_1 \leq \delta \qquad \forall \;\; n \geq N\,. \tag{8.31}$$

Since δ and r are arbitrary, Eq. (8.31) shows that $D_k(\gamma'') \geq (1-\varepsilon)D_k(\rho) + \varepsilon \operatorname{ld} d$, as suggested. This part does not rely on the constraint $D_2(\varrho) > 0$. ▲

The proof of the converse implication is based on Vidal's mixing lemma. Let us focus on $D_1(\varrho)$ first. Applying Lemma 8.11 and then Lemma 8.13, we see that for all sufficiently large integers $n \in \mathbb{N}$,

$$\begin{aligned} n\,D_1(\gamma'') &= D_1(\gamma''^{\otimes n}) \\ &\leq D_1\big((\varrho \otimes |0\rangle\langle 0|)^{\otimes\lfloor(1-\varepsilon)n\rfloor} \otimes (|\Omega_d\rangle\langle\Omega_d| \otimes |1\rangle\langle 1|)^{\otimes\lfloor\varepsilon n\rfloor}\big) \\ &\leq (1-\varepsilon)n\,D_1(\varrho) + \varepsilon\, n\, \operatorname{ld} d\,. \end{aligned} \tag{8.32}$$

Dividing by n, we obtain the desired result.

Since we do not yet know whether Lemma 8.13 holds for $D_2(\varrho)$, a different approach is needed in this case. We start by distilling maximally entangled qudit pairs from an ε-fraction of ϱ. For sufficiently large $n \in \mathbb{N}$,

$$\rho^{\otimes \lfloor \varepsilon n \rfloor} \mapsto |\Omega_d\rangle\langle\Omega_d|^{\otimes \lfloor \frac{D_2(\varrho)}{\operatorname{ld} d} \varepsilon n \rfloor}. \tag{8.33}$$

The resulting output is then mixed with ϱ according to Vidal's lemma,

$$\varrho^{\otimes \lfloor (1-\varepsilon)\frac{D_2(\rho)}{\operatorname{ld} d} n \rfloor} \otimes |\Omega_d\rangle\langle\Omega_d|^{\otimes \lfloor \frac{D_2(\varrho)}{\operatorname{ld} d} \varepsilon n \rfloor} \mapsto \gamma''^{\otimes \lfloor \frac{D_2(\varrho)}{\operatorname{ld} d} n \rfloor}. \tag{8.34}$$

Distilling the output state $\gamma''^{\otimes \lfloor \frac{D_2(\rho)}{\operatorname{ld} d} n \rfloor}$, in summary we have implemented the protocol

$$\varrho^{\otimes \lfloor n(\varepsilon + (1-\varepsilon)\frac{D_2(\varrho)}{\operatorname{ld} d}) \rfloor} \mapsto |\Omega_d\rangle\langle\Omega_d|^{\otimes \lfloor \frac{D_2(\varrho)}{\operatorname{ld} d} \frac{D_2(\gamma'')}{\operatorname{ld} d} n \rfloor}, \tag{8.35}$$

which corresponds to a singlet distillation rate

$$r := D_2(\varrho)\, \frac{D_2(\gamma'')}{(1-\varepsilon)D_2(\varrho) + \varepsilon \operatorname{ld} d}\,. \tag{8.36}$$

If $D_2(\varrho) > 0$ and $D_2(\gamma'') > (1-\varepsilon)D_2(\varrho) + \varepsilon \operatorname{ld} d$, then we would have given a distillation protocol for ϱ which achieves a rate $r > D_2(\varrho)$. By definition of $D_2(\varrho)$ this is impossible. This concludes the proof of Lemma 8.15. ■

The proof of the following lemma shows that γ' may be obtained from γ'' by 1-way LOCC operations alone:

Lemma 8.16. *With γ' and γ'' defined as in Eqs. (8.27) and (8.28) resp., we have*

$$D_k(\gamma') \leq D_k(\gamma'')\,. \tag{8.37}$$

Proof: We start out by expanding

$$\big((1-\varepsilon)|0\rangle\langle 0| \otimes \rho + \varepsilon |1\rangle\langle 1| \otimes \sigma\big)^{\otimes n} = \sum_{x \in \{0,1\}^n} (1-\varepsilon)^{|x|} \varepsilon^{n-|x|} |x\rangle\langle x| \otimes \rho^{\otimes |x|} \otimes \sigma^{\otimes (n-|x|)}\,, \tag{8.38}$$

where $|x|$ is the Hamming weight of the string x. Since the entanglement cost satisfies the upper bound $E_C(\sigma) \leq \operatorname{ld} d$, there is an LOCC protocol $(\Lambda_n)_{n\in\mathbb{N}}$ such that, for given $\delta > 0$, there is an $m_0 \equiv m_0(\delta)$ such that for all $m \geq m_0$

$$\|\Lambda_m(|\Omega_d\rangle\langle\Omega_d|^{\otimes m}) - \sigma^{\otimes m}\|_1 \leq \delta\,. \tag{8.39}$$

Since the input state is pure, a result due to Lo and Popescu [LP01] shows that this transformation may be implemented by 1-way LOCC operations alone (see also Th. 8.1 in Hayashi's text [Hay06^b]). For given δ, we easily see that there exists an $n_0 \equiv n_0(\delta)$ such that for all $n \geq n_0$

$$\Big\| \sum_{\substack{x \in \{0,1\}^n \\ |x| > (1-\varepsilon+\delta)n}} (1-\varepsilon)^{|x|} \varepsilon^{n-|x|} |x\rangle\langle x| \otimes \rho^{\otimes |x|} \otimes \sigma^{\otimes (n-|x|)} \Big\|_1 \leq \delta\,. \tag{8.40}$$

Note that the same inequality holds when $\sigma^{\otimes(n-|x|)}$ is replaced with $\Lambda_n(|\Omega_d\rangle\langle\Omega_d|^{\otimes(n-|x|)})$. For all n satisfying $n \geq n_0$ as well as $(\varepsilon-\delta)n \geq m_0$, we therefore find

$$\begin{aligned}
&|| \sum_{x\in\{0,1\}^n} (1-\varepsilon)^{|x|}\varepsilon^{n-|x|}|x\rangle\langle x| \otimes \rho^{\otimes|x|} \otimes \sigma^{\otimes(n-|x|)} \\
&\qquad - \sum_{x\in\{0,1\}^n} (1-\varepsilon)^{|x|}\varepsilon^{n-|x|}|x\rangle\langle x| \otimes \rho^{\otimes|x|} \otimes \Lambda_n(|\Omega_d\rangle\langle\Omega_d|^{\otimes(n-|x|)})||_1 \\
&\leq || \sum_{\substack{x\in\{0,1\}^n \\ |x|\leq(1-\varepsilon+\delta)n}} (1-\varepsilon)^{|x|}\varepsilon^{n-|x|}|x\rangle\langle x| \otimes \rho^{\otimes|x|} \otimes \sigma^{\otimes(n-|x|)} \\
&\qquad - \sum_{\substack{x\in\{0,1\}^n \\ |x|\leq(1-\varepsilon+\delta)n}} (1-\varepsilon)^{|x|}\varepsilon^{n-|x|}|x\rangle\langle x| \otimes \rho^{\otimes|x|} \otimes \Lambda_n(|\Omega_d\rangle\langle\Omega_d|^{\otimes(n-|x|)})||_1 \\
&\qquad + \| \sum_{\substack{x\in\{0,1\}^n \\ |x|>(1-\varepsilon+\delta)n}} (1-\varepsilon)^{|x|}\varepsilon^{n-|x|}|x\rangle\langle x| \otimes \rho^{\otimes|x|} \otimes \sigma^{\otimes(n-|x|)}\|_1 \\
&\qquad + \| \sum_{\substack{x\in\{0,1\}^n \\ |x|>(1-\varepsilon+\delta)n}} (1-\varepsilon)^{|x|}\varepsilon^{n-|x|}|x\rangle\langle x| \otimes \rho^{\otimes|x|} \otimes \Lambda_n(|\Omega_d\rangle\langle\Omega_d|^{\otimes(n-|x|)})\|_1 \\
&\leq 3\,\delta\,.
\end{aligned} \tag{8.41}$$

In summary, we have shown that for all $n \geq n_0$,

$$\|\gamma'^{\otimes n} - \Lambda_n(\gamma''^{\otimes n})\|_1 \leq 3\,\delta\,. \tag{8.42}$$

Therefore, given a protocol $(\Gamma_m)_{m\in\mathbb{N}}$ that distills singlets from γ' at rate r, the protocol $(\Gamma_m \circ \Lambda_m)_{m\in\mathbb{N}}$ will distill singlets from γ'' at the same rate. ■

We now have all the ingredients for the

Proof of Prop. 8.10: Since γ can be obtained from γ' by tracing out the flag system, it is evident that $D_k(\gamma) \leq D_k(\gamma')$. Invoking Lemma 8.16 and Lemma 8.15, we can then conclude that

$$D_k(\gamma) \leq D_k(\gamma') \leq D_k(\gamma'') = (1-\varepsilon)D_k(\rho) + \varepsilon \operatorname{ld} d\,, \tag{8.43}$$

provided that $D_2(\varrho) > 0$, which is the desired result. ■

8.5 Boundary

Prop. 8.10 applies to states with non-vanishing 2-way distillable entanglement only. A different approach is thus needed to show continuity at the boundary of the set $\mathcal{D}_2$. We will present two different (though related) proofs in this Section. One of them is based on hypothesis testing and applies to convex mixtures of the form investigated

in Sec. 8.4. The other one relies on a generalized version of the relative entropy of entanglement and applies to all states in the neighborhood of the non-distillable states. Both approaches are limited to the 2-way distillable entanglement, and we will point out where 2-way classical communication is required.

8.5.1 A Bound Based on Hypothesis Testing

The following Proposition extends Prop. 8.10 to states at the boundary of the set $\mathcal{D}_2$.

Proposition 8.17. *If $D_2(\varrho) = 0$ and $\gamma = (1-\varepsilon)\rho + \varepsilon\sigma$, then*

$$D_2(\gamma) \leq 6\,\varepsilon \operatorname{ld} d + 2\,\varepsilon + \eta(\varepsilon) + \eta(2\varepsilon)\,, \tag{8.44}$$

where $\eta(s) := -s \operatorname{ld} s$ for $s \in (0,1]$.

The proof of Prop. 8.17 relies on the following upper bound on the 2-way distillable entanglement, which appears as Th. 8.2 in Hayashi's text [Hay06^b]:

Lemma 8.18. *For any pair of states $\tau, \sigma \in \mathcal{B}_*(\mathcal{H} \otimes \mathcal{H})$ such that $D_2(\sigma) = 0$ we have*

$$D_2(\tau) \leq H(\tau \parallel \sigma)\,, \tag{8.45}$$

where $H(\tau \parallel \sigma) = \operatorname{tr} \tau \operatorname{ld} \tau - \operatorname{tr} \tau \operatorname{ld} \sigma$ denotes the quantum relative entropy (cf. App. A).

The proof of Lemma 8.18, as it appears in [Hay06^b], is based on hypothesis testing. In Hayashi's text, the reference state σ is assumed to be separable. However, the only assumption on σ in that proof is that it cannot be transformed into a state with high singlet fraction $F := \langle \Omega_d | \sigma | \Omega_d \rangle$ by LOCC-operations alone, and this applies to non-distillable states as well:

Lemma 8.19. *For any state $\sigma \in \mathcal{B}_*(\mathcal{H} \otimes \mathcal{H})$ with $d := \dim \mathcal{H}$ and $D_2(\sigma) = 0$ we have*

$$\langle \Omega_d | \sigma | \Omega_d \rangle \leq \frac{1}{d}\,. \tag{8.46}$$

The proof of Lemma 8.19 is due to M. and P. Horodecki [HH99], who proposed an explicit distillation protocol that yields a positive rate for any state σ with singlet fraction $\langle \Omega_d | \sigma | \Omega_d \rangle > 1/d$. In essence, their protocol goes as follows: Twirling, a 1-way LOCC operation, will transform the state σ into an isotropic state while preserving the singlet fraction. A local projection then yields a two-qubit isotropic state with singlet fraction $F > 1/2$. The 2-way purification protocol proposed by Bennett *et al.* [BBP$^+$96^a] will turn a collection of such states into a smaller number of states with higher singlet fraction. This process is iterated until the singlet fraction F is sufficiently high for the hashing protocol [BDS$^+$96] to take over, which will attain the limit $F \to 1$ for some positive distillation rate.

The hashing protocol again requires 1-way LOCC operations only, so the purification protocol is the only part in which backward communication is needed. If there were a 1-way protocol, Lemma 8.18 would likewise apply to 1-way distillable entanglement. However, a simple example shows that in general no such 1-way protocol exists: the two-qubit Werner state

$$\sigma_{\frac{5}{8}} := \frac{1}{2}|\Omega\rangle\langle\Omega| + \frac{1}{2}\,\frac{\mathbb{1}}{4} \tag{8.47}$$

has singlet fraction $F = 5/8 > 1/2$, and hence is 2-way distillable according to Lemma 8.19. However, Bennett *et al.* [BDS$^+$96] have shown that $D_1(\sigma_{\frac{5}{8}}) = 0$. In fact, all Werner states with singlet fraction $F < 3/4$ cannot be distilled by 1-way LOCC operations alone [BDS$^+$96, KL97].

For the proof of Prop. 8.17 we need one more simple Lemma on the distillability of mixtures:

Lemma 8.20. *Let $\sigma, \tau \in \mathcal{B}_*(\mathcal{H}\otimes\mathcal{H})$ be quantum states, and assume that τ is separable. Then for any $p \in [0,1]$,*

$$D_k\big((1-p)\,\sigma + p\,\tau\big) \leq D_k(\sigma)\,. \tag{8.48}$$

Proof of Lemma 8.20: Clearly, a tensor product state $\sigma \otimes \tau$ can be transformed to $(1-p)\,\sigma + p\,\tau$ by 1-way LOCC operations for any two states $\sigma, \tau \in \mathcal{B}_*(\mathcal{H}\otimes\mathcal{H})$, and any $p \in [0,1]$. Hence,

$$D_k\big((1-p)\,\sigma + p\,\tau\big) \leq D_k(\sigma\otimes\tau)\,. \tag{8.49}$$

If τ is separable, then $D_k(\sigma\otimes\tau) = D_k(\sigma)$, concluding the proof. ■

In fact, Vidal's asymptotic mixing (Lemma 8.11) improves on this bound,

$$D_k\big((1-p)\,\sigma + p\,\tau\big) \leq (1-p)\,D_k(\sigma) \tag{8.50}$$

for any $p \in [0,1]$ and separable state τ. But we will not need this stronger bound in the

Proof of Prop. 8.17: The techniques we use are inspired by Donald and Horodecki [DH01]. For $\varrho, \sigma \in \mathcal{B}_*(\mathcal{H}\otimes\mathcal{H})$ and $\varepsilon > 0$, we again set $\gamma := (1-\varepsilon)\varrho + \varepsilon\,\sigma$. Clearly, for any positive operator X and any positive constant c we have

$$\operatorname{ld}(X + c)\,\mathbb{1} \geq \max\big\{\operatorname{ld} X,\, \operatorname{ld} c\big\}\,. \tag{8.51}$$

We may now bound the relative entropy as follows:

$$\begin{aligned} H(\gamma \parallel (1-\varepsilon)\varrho + \varepsilon\,\frac{\mathbb{1}}{d^2}) &= -H(\gamma) - \operatorname{tr}(1-\varepsilon)\varrho\operatorname{ld}\big((1-\varepsilon)\varrho + \varepsilon\frac{\mathbb{1}}{d^2}\big) \\ &\qquad - \varepsilon\operatorname{tr}\sigma\operatorname{ld}\big((1-\varepsilon)\varrho + \varepsilon\frac{\mathbb{1}}{d^2}\big) \\ &\leq -H(\gamma) - \operatorname{tr}\varrho\operatorname{ld}(1-\varepsilon)\varrho - \varepsilon\operatorname{tr}\sigma\operatorname{ld}\varepsilon\,\frac{\mathbb{1}}{d^2} \\ &= -H(\gamma) + H(\varrho) - \operatorname{ld}(1-\varepsilon) - \varepsilon\operatorname{ld}\varepsilon + 2\,\varepsilon\operatorname{ld} d\,. \end{aligned} \tag{8.52}$$

By Fannes' inequality (cf. Prop. A.6), for sufficiently small ε we have

$$|H(\varrho) - H(\gamma)| \leq \|\varrho - \gamma\|_1 \operatorname{ld} d^2 + \eta(\|\varrho - \gamma\|_1) \leq 4\,\varepsilon \operatorname{ld} d + \eta(2\varepsilon)\,, \tag{8.53}$$

where $\eta(\varepsilon) = -\varepsilon \operatorname{ld} \varepsilon$. We then see from Eq. (8.52) that

$$H(\gamma \parallel (1-\varepsilon)\varrho + \varepsilon \frac{\mathbb{1}}{d^2}) \leq 6\,\varepsilon \operatorname{ld} d + 2\,\varepsilon + \eta(\varepsilon) + \eta(2\varepsilon)\,. \tag{8.54}$$

If $D_2(\varrho) = 0$, the state $(1-\varepsilon)\varrho + \varepsilon \frac{\mathbb{1}}{d^2}$ is likewise non-distillable due to Lemma 8.20, since $\frac{\mathbb{1}}{d^2}$ is separable. Lemma 8.18 now immediately yields the desired bound. ■

8.5.2 Accurate Entanglement

We will now give a second proof for continuity at the boundary of the set $\mathcal{D}_2$, which is based on some of the same techniques that we have used in Sec. 8.5.1, but does not rely on hypothesis testing. Moreover, it is not restricted to convex mixtures, but applies to any state in the neighborhood of the boundary of the non-distillable states. However, the result is again limited to 2-way distillable entanglement.

Proposition 8.21. *For any two quantum states $\varrho, \sigma \in \mathcal{B}_*(\mathcal{H} \otimes \mathcal{H})$ with $D_2(\sigma) = 0$ and $\|\varrho - \sigma\|_1 \leq 1/3$ we have*

$$D_2(\varrho) \leq 4\,\|\varrho - \sigma\|_1 \operatorname{ld} d + 4\,\|\varrho - \sigma\|_1 + 2\,\eta(\|\varrho - \sigma\|_1)\,, \tag{8.55}$$

where $d := \dim \mathcal{H}$ and $\eta(s) := -s \operatorname{ld} s$ for $s \in (0, 1]$.

Our strategy for the proof of Prop. 8.21 is to introduce a continuous entanglement measure which upper bounds D_2 and vanishes on the non-distillable states.

Definition 8.22. (Accurate Entanglement)
Let $\rho \in \mathcal{B}_*(\mathcal{H} \otimes \mathcal{H})$ be a bipartite quantum state. The k-way accurate entanglement *of ρ is given by*

$$A_k(\rho) = \inf_{\sigma \in \mathcal{D}_k^c} H(\rho||\sigma)\,, \tag{8.56}$$

where $\mathcal{D}_k^c = \{\sigma \in \mathcal{B}_*(\mathcal{H} \otimes \mathcal{H}) \mid D_k(\sigma) = 0\}$ denotes the complement of the set $\mathcal{D}_k$.

We know from Cor. 8.8 that the set of non-distillable quantum states $\mathcal{D}_k^c$ is closed, and hence compact. Since the relative entropy is lower semi-continuous [OP93], the infimum in Eq. (8.56) is attained [RS80].

The accurate entanglement A_k is a variant of the *relative entropy of entanglement* [VPR+97] (in which the infimum in Eq. (8.56) is taken over all separable states instead) and shares many of its properties, even though the set $\mathcal{D}_k^c$ may not be convex. It is evident from the definition that $A_k(\varrho)$ vanishes iff $\varrho \in \mathcal{D}_k^c$. We will show below that A_k is *asymptotically continuous* (which by definition just means that Eq. (8.60) holds),

and that $A_2(\varrho) \geq D_2(\varrho)$ for every quantum state ϱ. This will then imply that D_2 is continuous on the boundary of the non-distillable states.

According to Lemma 8.20, the set of non-distillable states $\mathcal{D}_k^c$ is closed under convex mixtures with separable states. We may thus apply the results of Donald and Horodecki [DH01] to conclude that A_k is asymptotically continuous:

Lemma 8.23. (Asymptotic Continuity of Accurate Entanglement)
For any pair of quantum states $\varrho, \sigma \in \mathcal{B}_*(\mathcal{H} \otimes \mathcal{H})$ such that $\|\varrho - \sigma\|_1 \leq 1/3$ we have:

$$|A_k(\varrho) - A_k(\sigma)| \leq 4\,\|\varrho - \sigma\|_1 \operatorname{ld} d + 4\,\|\varrho - \sigma\|_1 + 2\,\eta(\|\varrho - \sigma\|_1)\,, \tag{8.57}$$

where $d := \dim \mathcal{H}$, and again $\eta(s) := -s \operatorname{ld} s$ for $s \in (0, 1]$.

In addition, A_k cannot increase under k-way LOCC operations and is subadditive on tensor products. Moreover, $A_2(|\Omega_d\rangle\langle\Omega_d|) = \operatorname{ld} d$. We do not yet know whether $A_1(|\Omega_d\rangle\langle\Omega_d|)$ might exceed $\operatorname{ld} d$. We summarize what is known about the accurate entanglement in the following

Proposition 8.24. (Properties of Accurate Entanglement)
The accurate entanglement *A_k, as defined in Def. 8.22 above, has the following properties:*

(*i*) *A_k does not increase under k-way LOCC operations.*

(*ii*) *A_k is subadditive on tensor powers,*

$$A_k(\varrho^{\otimes n}) \leq n\, A_k(\varrho) \tag{8.58}$$

for any state $\varrho \in \mathcal{B}_*(\mathcal{H} \otimes \mathcal{H})$.

(*iii*) *A_k is asymptotically continuous.*

(*iv*) *A_2 is normalized,*

$$A_2(|\Omega_d\rangle\langle\Omega_d|) = \operatorname{ld} d\,. \tag{8.59}$$

Proof: Asymptotic continuity follows from Lemma 8.23,

$$\frac{|A_k(\varrho_n) - A_k(\sigma_n)|}{\operatorname{ld}\ \dim \mathcal{H}_n} \to 0 \qquad \text{for} \qquad \|\varrho_n - \sigma_n\|_1 \to 0 \tag{8.60}$$

for all sequences $(\varrho_n)_{n\in\mathbb{N}}, (\sigma_n)_{n\in\mathbb{N}}$ of quantum states on $\mathcal{H}_n := \mathcal{H}^{\otimes n} \otimes \mathcal{H}^{\otimes n}$.

In order to show (*i*), assume that Λ^k is a k-way LOCC operation. Since the set $\mathcal{D}_k^c$ is closed under Λ^k, and the quantum relative entropy does not increase under quantum operations (cf. App. A), we see that

$$A_k(\Lambda^k(\varrho)) = \inf_{\sigma\in\mathcal{D}_k^c} H(\Lambda^k(\varrho) \parallel \sigma) \leq H(\Lambda^k(\varrho) \parallel \Lambda^k(\tilde{\sigma})) \leq H(\varrho \parallel \tilde{\sigma}) = A_k(\varrho)\,, \tag{8.61}$$

where we have chosen $\tilde{\sigma} \in \mathcal{D}_k^c$ as the state that attains the infimum in the definition of $A_k(\varrho)$.

For the subadditivity under tensor powers, note that $H(\varrho^{\otimes n} \parallel \sigma^{\otimes n}) = n\, H(\varrho \parallel \sigma)$ holds for all states $\varrho, \sigma \in \mathcal{B}_*(\mathcal{H} \otimes \mathcal{H})$. Since $\mathcal{D}_k^c$ is closed under tensor product powers, we immediately see that

$$A_k(\varrho^{\otimes n}) = \inf_{\sigma \in \mathcal{D}_k^c} H(\varrho^{\otimes n} \parallel \sigma) \leq H(\varrho^{\otimes n} \parallel \tilde{\sigma}^{\otimes n}) = n\, H(\varrho \parallel \tilde{\sigma}) = A_k(\varrho)\,, \tag{8.62}$$

where again we have chosen $\tilde{\sigma} \in \mathcal{D}_k^c$ such that $A_k(\varrho) = H(\varrho \parallel \tilde{\sigma})$. Eq. (8.62) holds for every positive integer $n \in N$, which is the desired result.

We will now show the normalization of A_2 on maximally entangled states. For any Hermitian operator $X \in \mathcal{M}_d$ and any pure state $|\psi\rangle \in \mathbb{C}^d$ we have

$$f(\langle\psi|X|\psi\rangle) \leq \langle\psi|f(X)|\psi\rangle \tag{8.63}$$

for every convex function f. The proof of Eq. (8.63) is straightforward: assume that $X = \sum_i x_i\, |i\rangle\langle i|$ is the spectral decomposition for the operator X, then $p_i := |\langle\psi|i\rangle|^2$ defines a probability distribution. By convexity of the function f, we thus have

$$f(\langle\psi|X|\psi\rangle) = f\big(\sum_i |\langle\psi|i\rangle|^2\, x_i\big) \leq \sum_i |\langle\psi|i\rangle|^2\, f(x_i) = \langle\psi|f(X)|\psi\rangle\,, \tag{8.64}$$

as suggested. Since the function $f(s) := -\mathrm{ld}\, s$ is convex, an application of Eq. (8.63) immediately shows that

$$H(|\Omega_d\rangle\langle\Omega_d| \parallel \sigma) = -\langle\Omega_d|\mathrm{ld}\,\sigma|\Omega_d\rangle \geq -\mathrm{ld}\,\langle\Omega_d|\sigma|\Omega_d\rangle \tag{8.65}$$

for all quantum states $\sigma \in \mathcal{B}_*(\mathcal{H} \otimes \mathcal{H})$. If $D_2(\sigma) = 0$, then Lemma 8.19 implies that

$$A_2(|\Omega_d\rangle\langle\Omega_d|) = \inf_{\sigma \in \mathcal{D}_k^c} H(|\Omega_d\rangle\langle\Omega_d| \parallel \sigma) \geq \mathrm{ld}\, d\,. \tag{8.66}$$

Vedral and Plenio [VP98] have shown that the relative entropy of entanglement coincides with the von Neumann reduced entropy on every pure state, which implies that there exists a separable state σ which attains the bound in Eq. (8.66). This completes the proof. ■

By the results of Donald *et al.* [DHR02], the properties established in Prop. 8.24 are sufficient to guarantee that A_2 upper bounds the 2-way distillable entanglement, circumventing Hayashi's hypothesis testing argument:

Corollary 8.25. *For every quantum state $\varrho \in \mathcal{B}_*(\mathcal{H} \otimes \mathcal{H})$,*

$$D_2(\varrho) \leq A_2(\varrho)\,. \tag{8.67}$$

Proof: The proof can be found in Sec. VII of [DHR02]; we reproduce it here for completeness. Let $\varrho \in \mathcal{B}_*(\mathcal{H} \otimes \mathcal{H})$ be a quantum state and $\varepsilon > 0$. Then by the definition of 2-way distillable entanglement, there exist an LOCC protocol $(\Lambda_n^2)_{n\in\mathbb{N}}$ and a positive integer $N \in \mathbb{N}$ such that

$$|D_2(\varrho) - \frac{\mathrm{ld}\, d_n}{n}| \leq \varepsilon \qquad \forall \quad n \geq N \tag{8.68}$$

and $\lim_{n\to\infty} \| |\Omega_{d_n}\rangle\langle\Omega_{d_n}| - \Lambda_n^2(\varrho^{\otimes n})\|_1 = 0$. By the asymptotic continuity of A_2, N may be chosen large enough such that

$$\frac{|A_2(|\Omega_{d_n}\rangle\langle\Omega_{d_n}|) - A_2(\Lambda_n^2(\varrho^{\otimes n}))|}{n} \leq \varepsilon \qquad \forall \quad n \geq N\,. \tag{8.69}$$

Using the subadditivity of A_2 on tensor product powers, the monotonicity under LOCC operations, and the normalization on the maximally entangled vector Ω_{d_n}, we have for all $n \geq N$:

$$A_2(\varrho) \geq \frac{A_2(\varrho^{\otimes n})}{n} \geq \frac{A_2(\Lambda_n^2(\varrho^{\otimes n}))}{n} \geq \frac{A_2(|\Omega_{d_n}\rangle\langle\Omega_{d_n}|)}{n} - \varepsilon = \frac{\mathrm{ld}\, d_n}{n} - \varepsilon \geq D_2(\varrho) - 2\,\varepsilon\,. \tag{8.70}$$

Since $\varepsilon > 0$ is arbitrary, the Corollary is proved. ■

From Lemma 8.23 and Corollary 8.25 we see that D_2 is upper bounded by a continuous function which vanishes on the boundary of the non-distillable states, and is hence likewise continuous. If $\varrho, \sigma \in \mathcal{B}_*(\mathcal{H} \otimes \mathcal{H})$ are quantum states such that $D_2(\sigma) = 0$ then

$$D_2(\varrho) \leq A_2(\varrho) \leq 4\,\|\varrho - \sigma\|_1 \,\mathrm{ld}\, d + 4\,\|\varrho - \sigma\|_1 + 2\,\eta(\|\varrho - \sigma\|_1)\,, \tag{8.71}$$

which proves Prop. 8.21.

8.6 Faithful States

Given a quantum state $\varrho \in \mathcal{B}_*(\mathcal{H}\otimes\mathcal{H})$ and $\gamma := (1-\varepsilon)\varrho + \varepsilon\sigma$ as in Sec. 8.4, we conclude from Prop. 8.10 that

$$D_1(\gamma) \leq D_1(\varrho) + \varepsilon\,\mathrm{ld}\, d\,. \tag{8.72}$$

For the 2-way distillable entanglement, we can combine Prop. 8.10 with the boundary results in Prop. 8.17 to give

$$D_2(\gamma) \leq D_1(\varrho) + 6\,\varepsilon\,\mathrm{ld}\, d + 2\,\varepsilon + \eta(\varepsilon) + \eta(2\varepsilon)\,. \tag{8.73}$$

Away from the boundary of the state space, these results may be inverted to give corresponding upper bounds on the distillable entanglement $D_k(\varrho)$ in terms of $D_k(\gamma)$: For $c \in (0,1)$, we define

$$F_c := \{\rho \in \mathcal{B}_*(\mathcal{H} \otimes \mathcal{H}) \;\mid\; \rho \geq c\mathbb{1}\}\,, \tag{8.74}$$

i. e., the convex set of *faithful states* containing a fraction of the identity. For faithful states, the bounds in Eqs. (8.72) and (8.73) can be inverted with the substitution $\varepsilon \mapsto \frac{\varepsilon}{c}$.

Proposition 8.26. *Let $c \in (0,1)$. Then for any faithful state $\varrho \in F_c \subset \mathcal{B}_*(\mathcal{H}\otimes\mathcal{H})$ and $\gamma = (1-\varepsilon)\varrho + \varepsilon\sigma$ for some $\varepsilon > 0$ and $\sigma \in \mathcal{B}_*(\mathcal{H}\otimes\mathcal{H})$ we have,*

$$D_1(\varrho) \leq D_1(\gamma) + \frac{\varepsilon}{c}\operatorname{ld} d \qquad \text{and} \tag{8.75}$$

$$D_2(\varrho) \leq D_1(\gamma) + 6\,\frac{\varepsilon}{c}\operatorname{ld} d + 2\,\frac{\varepsilon}{c} + \eta\Big(\frac{\varepsilon}{c}\Big) + \eta\Big(\frac{2\varepsilon}{c}\Big)\,. \tag{8.76}$$

Proof: Since $\varrho - c\,\sigma \geq \varrho - c\,\mathbb{1} \geq 0$, $\hat{\sigma} := \frac{\varrho - c\,\sigma}{1-c} \in \mathcal{B}_*(\mathcal{H}\otimes\mathcal{H})$ defines a quantum state. A straightforward calculation shows that ϱ can be given the expansion $\varrho = (1-\varepsilon')\gamma + \varepsilon'\hat{\sigma}$ with the mixing parameter

$$\varepsilon' := \frac{\varepsilon\big(\frac{1}{c}-1\big)}{1+\varepsilon\big(\frac{1}{c}-1\big)} \leq \frac{\varepsilon}{c} - \varepsilon \leq \frac{\varepsilon}{c}\,. \tag{8.77}$$

The desired result now immediately follows from Prop. 8.10 in conjunction with the boundary results in Prop. 8.17. ■

We can now apply Lemma 8.9 to conclude that both the 1-way and the 2-way distillable entanglement are asymptotically continuous on the set F_c of faithful states for any fixed $c \in (0,1)$.

Theorem 8.27. (Continuity on Faithful States)
Let $c \in (0,1)$. Then for any pair of faithful quantum states $\varrho, \tau \in F_c \subset \mathcal{B}_*(\mathcal{H}\otimes\mathcal{H})$ with $\|\varrho - \tau\|_1 \leq \varepsilon$ we have

$$|D_1(\varrho) - D_1(\tau)| \leq 2\,\frac{\varepsilon}{c}\operatorname{ld} d \qquad \text{and} \tag{8.78}$$

$$|D_2(\varrho) - D_2(\tau)| \leq 12\,\frac{\varepsilon}{c}\operatorname{ld} d + 4\,\frac{\varepsilon}{c} + 2\,\eta\Big(\frac{\varepsilon}{c}\Big) + 2\,\eta\Big(\frac{2\varepsilon}{c}\Big)\,, \tag{8.79}$$

where $d := \dim\mathcal{H}$.

This is the main result of this Chapter. We do not yet know how to do without faithfulness, though we conjecture that similar bounds apply globally.

8.7 Quantum Channel Capacity

Quantum channel capacity has been shown to be lower semi-continuous in [KW02]. Using the duality of states and quantum channels and the results on conclusive teleportation, we can prove that the quantum channel capacity Q_2 assisted by 2-way classical communication is continuous on the boundary of the quantum channels with vanishing capacity, with explicit bounds:

Proposition 8.28. *Let $T, S\colon \mathcal{B}(\mathcal{H}) \to \mathcal{B}(\mathcal{K})$ be a pair of quantum channels such that $\|T - S\|_{cb} \leq \varepsilon$ and $Q_2(S) = 0$. We then have*

$$Q_2(T) \leq d_{\mathcal{H}}^2\big(4\,\varepsilon\operatorname{ld} d_{\mathcal{K}} + 4\,\varepsilon + 2\,\eta(\varepsilon)\big)\,, \tag{8.80}$$

where $d_{\mathcal{H}} := \dim\mathcal{H}$, $d_{\mathcal{K}} := \dim\mathcal{K}$, and $\eta(s) := -s\operatorname{ld} s$ for $s \in (0,1]$.

The proof of Prop. 8.28 relies on the duality of channel capacity and distillable entanglement: Given a quantum channel $T\colon \mathcal{B}(\mathcal{H}) \to \mathcal{B}(\mathcal{K})$, we define the bipartite state

$$\varrho_T := (T_* \otimes \mathrm{id}\,)|\Omega_{\mathcal{K}}\rangle\langle\Omega_{\mathcal{K}}|\,, \tag{8.81}$$

where $|\Omega_{\mathcal{K}}\rangle$ is a maximally entangled state on $\mathcal{K} \otimes \mathcal{K}$. If Alice and Bob want to send information through the quantum channel T undistorted, a possible strategy [BDS+96] for them is to create the state ϱ_T by applying the channel T_* to half of a maximally entangled state, distill the state ϱ_T using a k-way LOCC protocol, and then use the resulting ebit pairs and classical forward communication to teleport information from Alice to Bob. Hence, the k-way capacity of the channel T is no smaller than the k-way distillable entanglement of the quantum state ϱ_T, $Q_k(T) \geq D_k(\varrho_T)$. Since the quantum channel capacity is unaffected by classical forward communication alone (cf. 5.4.2), we have $Q(T) \geq D_1(\varrho_T)$.

Conversely, using a bipartite quantum state ϱ as a resource in the standard teleportation protocol implements a quantum channel T_ϱ. A possible way [BDS+96] for Alice and Bob to distill the state ϱ is then to implement the quantum channel T_ϱ, so that Alice can send maximally entangled states to Bob. Hence, $D_k(\varrho) \geq Q_k(T_\varrho)$.

In general, this does not yield a one-to-one relation between distillable entanglement and quantum capacity, for there are states such that $\varrho_{T_\varrho} \neq \varrho$ [BDS+96]. However, for isotropic states equivalence does hold, and the corresponding channels T_ϱ are partially depolarizing [HHH99].

In the general case, so-called *conclusive teleportation* [BHM04] allows to implement the channel T at least probabilistically, resulting in the following

Lemma 8.29. *For any quantum channel $T\colon \mathcal{B}(\mathcal{H}) \to \mathcal{B}(\mathcal{K})$ we have*

$$\frac{Q_k(T)}{d_{\mathcal{H}}^2} \leq D_k(\varrho_T) \leq Q_k(T)\,. \tag{8.82}$$

A proof of Lemma 8.29, along the lines sketched above, can be found in [HN05].

If $Q_k(T) \approx 0$, the bound in Eq. (8.82) becomes sharp. The proof of Prop. 8.28 is then straightforward: Assuming $Q_2(S) = 0$, we conclude from Eq. (8.82) that $D_2(\varrho_S) = 0$. Moreover,

$$\|\varrho_T - \varrho_S\|_1 = \|\big((T_* - S_*) \otimes \mathrm{id}\,\big)|\Omega_{\mathcal{K}}\rangle\langle\Omega_{\mathcal{K}}|\|_1 \leq \|T - S\|_{cb} \leq \varepsilon\,. \tag{8.83}$$

The proof now follows from combining the left half of Eq. (8.82) with Eq. (8.55).

8.8 Summary and Outlook

In this Chapter we have investigated the continuity of distillable entanglement and we have shown uniform norm bounds for faithful states (cf. Th. 8.27). Since we also

know from Prop. 8.5 that both the 1-way and the 2-way distillable entanglement are continuous near pure states, the only states for which continuity bounds are presently missing are the non-pure states on the boundary of the state space. Given such a state ϱ, it would be enough to find an upper bound on $D_k(\varrho)$ in terms of $D_k(\gamma)$ for the convex mixture $\gamma := (1-\varepsilon)\varrho + \frac{1}{d^2}\mathbb{1}$ and (small) $\varepsilon > 0$. A corresponding lower bound, $D_k(\gamma) \leq D_k(\varrho)$, is immediate from Lemma 8.20, since $\frac{1}{d^2}\mathbb{1}$ is separable. Because such a γ is faithful, the general result would then immediately follow from Th. 8.27, in conjunction with the triangle inequality.

The lower semi-continuity of the distillable entanglement, as discussed in Sec. 8.3, does yield such an upper bound on $D_k(\varrho)$ in terms of $D_k(\gamma)$, but these bounds are in general dependent on the quantum state ϱ under consideration, and hence do not go together well with Lemma 8.9, which requires uniform bounds.

Our uniform norm bounds suggest that distillable entanglement can be given an unambiguous operational interpretation. However, for the working physicist it is probably not enough to know that two states ϱ and σ that are almost indistinguishable contain similar amounts of distillable entanglement. He would also want to be ascertained that there exists a distillation protocol that attains the maximal rate and works well for *both* ϱ and σ. Such *universal* distillation protocols have been found for pure states [MH05], in a slightly different scenario. Despite its considerable practical and operational relevance, the general case remains very much open.

These questions all apply to quantum channel capacities as well. In Sec. 8.7 we have derived continuity bounds for the quantum capacity Q_2 assisted by 2-way classical side channels — yet only in the neighborhood of channels with vanishing capacity. It seems highly desirable to extend these continuity bounds both to general channels away from the boundary, and to the other classical and quantum channel capacities discussed in Ch. 5. We know from Sec. 5.5.2 that the entanglement-assisted capacities $C^E = 2\,Q^E$ can be expressed in terms of a single-letter entropic quantity, and hence continuity immediately follows from Fannes' inequality (cf. Prop. A.6). But little is known so far about the continuity properties of all the other channel capacities, leaving plenty of room for future research.

Appendix A

Entropic Information Measures and Their Basic Properties

Due to the equipartition of eigenvalues, entropic quantities play a major role as information measures in asymptotic information theory, both classical and quantum (cf. Sec. 7.1 for further details). For easy reference, in this Section we collect some of the quantities that have particularly wide currency, and list some of their basic properties. All these results are well-known in the quantum information community and may be found in the standard textbooks on the subject [NC00, Hay06^b]. Most of the theory applies to infinite-dimensional Hilbert spaces as well, but our presentation will be limited to the finite-dimensional setup.

A.1 Relative Entropy

A convenient starting point for this Appendix is the *quantum relative entropy*, from which all other important information measures (both classical and quantum) can be derived. For two positive operators $X, Y \in \mathcal{B}(\mathcal{H})$ (not necessarily trace-normalized) on some finite-dimensional Hilbert space $\mathcal{H}$ we define the *relative entropy* $H(X \parallel Y)$ of X to Y as

$$H(X \parallel Y) := \begin{cases} \operatorname{tr} X(\operatorname{ld} X - \operatorname{ld} Y) & : \quad \operatorname{supp}(X) \subset \operatorname{supp}(Y)\,, \\ \infty & : \quad \text{else.} \end{cases} \tag{A.1}$$

Though not a metric in the mathematical sense (both symmetry and the triangle inequality will in general fail to hold), we will see below that the relative entropy does have some distance-like properties, which make it a convenient entropic measure of the closeness of two quantum states. It provides an upper bound on the trace norm distance:

Proposition A.1. (Trace Norm Bound)
For any two positive operators $X, Y \in \mathcal{B}(\mathcal{H})$ we have

$$H(X \parallel Y) \geq \frac{\mathrm{ld}\, e}{2} \|X - Y\|_1^2 + \mathrm{ld}\,(e)\,(\mathrm{tr}\, Y - \mathrm{tr}\, X)\,. \tag{A.2}$$

In particular, Prop. A.1 implies that for any two quantum states $\varrho, \sigma \in \mathcal{B}_*(\mathcal{H})$

$$H(\varrho \parallel \sigma) \geq \frac{\mathrm{ld}\, e}{2} \|\varrho - \sigma\|_1^2 \geq 0\,, \tag{A.3}$$

with equality iff $\varrho = \sigma$, a result sometimes known as *Klein's inequality* [Kle31]. Lindblad [Lin75] and Uhlmann [Uhl77] were the first to show that the relative entropy cannot increase under quantum operations:

Proposition A.2. (Monotonicity of the Relative Entropy)
For any quantum channel $T\colon \mathcal{B}(\mathcal{K}) \to \mathcal{B}(\mathcal{H})$ and any two positive operators $X, Y \in \mathcal{B}(\mathcal{H})$ we have

$$H(T_*(X) \parallel T_*(Y)) \leq H(X \parallel Y)\,. \tag{A.4}$$

A.2 Entropy and Related Information Measures

The von Neumann entropy of a quantum state $\varrho \in \mathcal{B}_*(\mathcal{H})$ (cf. Sec. 2.8) is recovered from the quantum relative entropy simply by choosing $Y = \mathbb{1}$ as reference in Eq. (A.1): $H(\varrho) = -H(\varrho \parallel \mathbb{1}) = -\mathrm{tr}\, \varrho\, \mathrm{ld}\, \varrho$. Setting $\sigma := \mathrm{tr}_B\, \varrho \otimes \mathrm{tr}_A\, \varrho$ for some bipartite quantum state $\varrho \in \mathcal{B}_*(\mathcal{H}_A \otimes \mathcal{H}_B)$, an application of Klein's inequality shows that the von Neumann entropy is *subadditive* and satisfies a triangle inequality,

$$|H(\mathrm{tr}_A \varrho) - H(\mathrm{tr}_B \varrho)| \leq H(\varrho) \leq H(\mathrm{tr}_A \varrho) + H(\mathrm{tr}_B \varrho)\,. \tag{A.5}$$

The equality condition in Klein's inequality implies that equality holds on the RHS of Eq. (A.5) iff ϱ is a product state, $\varrho = \mathrm{tr}_B\, \varrho \otimes \mathrm{tr}_A\, \varrho$. A direct calculation leads to the so-called *joint entropy theorem* for classical-quantum states:

Proposition A.3. (Joint Entropy Theorem)
Suppose $\{p_i, \varrho_i\}_{i=1}^I$ is a quantum ensemble with a classical probability distribution $\{p_i\}_{i=1}^I$ and quantum states $\varrho_i \in \mathcal{B}_*(\mathcal{H})$. Introducing an orthonormal basis $\{|i\rangle\}_{i=1}^I$ for the quantum system $\mathbb{C}^I$, we have

$$H\Big(\sum_{i=1}^I p_i\, |i\rangle\langle i| \otimes \varrho_i\Big) = \sum_{i=1}^I p_i\, H(\varrho_i) + H(\{p_i\})\,, \tag{A.6}$$

where $H(\{p_i\})$ denotes the Shannon entropy of the classical distribution $\{p_i\}_{i=1}^I$, as in Eq. (2.53).

The subadditivity inequality and the joint entropy theorem can be applied to show that the von Neumann entropy is concave, yet not too much:

Proposition A.4. (Concavity of the Entropy)
For any ensemble $\{p_i, \varrho_i\}_{i=1}^I$ with a classical probability distribution $\{p_i\}_{i=1}^I$ and quantum states $\varrho_i \in \mathcal{B}_*(\mathcal{H})$ we have

$$\sum_{i=1}^I p_i\, H(\varrho_i) \;\leq\; H(\varrho) \;\leq\; \sum_{i=1}^I p_i\, H(\varrho_i) + H(\{p_i\})\,, \tag{A.7}$$

where $\varrho := \sum_{i=1}^I p_i\, \varrho_i$ denotes the averaged state, and $H(\{p_i\})$ is again the Shannon entropy of the distribution $\{p_i\}_{i=1}^I$.

The *mutual information* $H(A:B)$ of a bipartite system in the state $\varrho \in \mathcal{B}_*(\mathcal{H}_A \otimes \mathcal{H}_B)$ with local restrictions $\varrho^A := \mathrm{tr}_B\, \varrho$ and $\varrho^B := \mathrm{tr}_A\, \varrho$ is given by

$$H(A:B) \equiv H(\varrho^A : \varrho^B) := H(\varrho \parallel \varrho^A \otimes \varrho^B) = H(\varrho^A) + H(\varrho^B) - H(\varrho)\,, \tag{A.8}$$

which is always non-negative due to the subadditivity of the von Neumann entropy. For classical-quantum states of the form $\varrho = \sum_i p_i\, |i\rangle\langle i| \otimes \varrho_i$ for an ensemble of quantum states $\varrho_i \in \mathcal{B}_*(\mathcal{H}_B)$, we immediately conclude from Prop. A.3 that

$$H(\varrho^A : \varrho^B) = H\Big(\sum_{i=1}^I p_i\, \varrho_i\Big) - \sum_{i=1}^I p_i\, H(\varrho_i) =: \chi\big(\{p_i, \varrho_i\}\big)\,, \tag{A.9}$$

the so-called *Holevo bound* [Hol73] of the quantum ensemble $\{p_i, \varrho_i\}$. If the quantum ensemble $\{p_i, T_*(\varrho_i)\}$ arises from sending the ensemble $\{p_i, \varrho_i\}$ through the quantum channel T_*, we usually write $\chi(T_*, \{p_i, \varrho_i\}) := \chi(\{p_i, T_*(\varrho_i)\})$ — a quantity that plays an eminent role as the asymptotic rate function for the transmission of classical information through quantum channels (cf. Sec 5.5.1).

For quantum information transfer, the corresponding rate function is the *coherent information*, defined in terms of the *conditional entropy* $H(A|B)$ of the bipartite state $\varrho \in \mathcal{B}_*(\mathcal{H}_A \otimes \mathcal{H}_B)$,

$$H(A|B) \equiv H(\varrho^A|\varrho^B) := -H(\varrho \parallel \mathbb{1} \otimes \varrho^B) = H(\varrho) - H(\varrho^B) = H(\varrho^A) - H(A:B)\,. \tag{A.10}$$

Given a quantum channel $T_*\colon \mathcal{B}_*(\mathcal{H}_{A'}) \to \mathcal{B}_*(\mathcal{H}_B)$ with input state $\varrho \in \mathcal{B}_*(\mathcal{H}_{A'})$, the *coherent information* $I_c(T_*, \varrho)$ is now defined [SN96, Llo97] as

$$I_c(T_*, \varrho) := H(T_*(\varrho)) - H(T_* \otimes \mathrm{id}\, |\psi_\varrho\rangle\langle\psi_\varrho|) = H(B) - H(AB) = -H(A|B)\,, \tag{A.11}$$

where $|\psi_\varrho\rangle \in \mathcal{H}_{A'} \otimes \mathcal{H}_A$ denotes a purification of the quantum state ϱ (cf. Sec. 2.6).

Since both the Holevo bound and the coherent information may be expressed in terms of the quantum relative entropy, we can apply Prop. A.2 to conclude that neither increases under quantum operations:

Proposition A.5. (Data Processing Inequalities)
Let $T_*\colon \mathcal{B}_*(\mathcal{H}_A) \to \mathcal{B}_*(\mathcal{H}_B)$ and $S_*\colon \mathcal{B}_*(\mathcal{H}_B) \to \mathcal{B}_*(\mathcal{H}_C)$ be quantum channels. For any quantum state $\varrho \in \mathcal{B}_*(\mathcal{H}_A)$ we then have

$$I_c(S_* \circ T_*, \varrho) \leq I_c(T_*, \varrho) \leq H(\varrho)\,. \tag{A.12}$$

Correspondingly, for any ensemble $\{p_i, \varrho_i\}$ with quantum states $\varrho_i \in \mathcal{B}_*(\mathcal{H}_A)$ we have

$$\chi\big(S_* \circ T_*, \{p_i, \varrho_i\}\big) \leq \chi\big(T_*, \{p_i, \varrho_i\}\big) \leq \chi\big(\{p_i, \varrho_i\}\big)\,. \tag{A.13}$$

A.3 Continuity Properties

M. Fannes [Fan73] has shown that the von Neumann entropy is continuous in trace norm topology:

Proposition A.6. (Fannes' Inequality)
For any two quantum states $\varrho, \sigma \in \mathcal{B}_*(\mathcal{H})$ such that $\|\varrho - \sigma\|_1 \leq 1/e$ we have

$$|H(\varrho) - H(\sigma)| \leq \|\varrho - \sigma\|_1 \operatorname{ld} d + \eta(\|\varrho - \sigma\|_1)\,, \tag{A.14}$$

where $\eta(s) := -s \operatorname{ld} s$ for $s > 0$, and $d := \dim \mathcal{H}$. Removing the restriction on the trace norm distance, we can derive the weaker bound

$$|H(\varrho) - H(\sigma)| \leq \|\varrho - \sigma\|_1 \operatorname{ld} d + \frac{1}{e \ln 2}\,. \tag{A.15}$$

Koenraad Audenaert has recently improved on these results with a continuity bound that can be saturated [Aud06]. However, the functional dependence on the trace norm distance and the dimension of the underlying Hilbert space is very similar to Fannes' inequality, so for most practical purposes these two bounds can be applied interchangeably. The continuity of the conditional information $H(\varrho^A|\varrho^B) := H(\varrho) - H(\varrho^B)$ can be derived from Fannes' or Audenaert's inequlity. However, a stronger bound by Alicki and Fannes [AF04] avoids the dependence on the dimension of the Hilbert space $\mathcal{H}_B$.

Proposition A.7. (Continuity of Quantum Conditional Information)
Let $\varrho, \sigma \in \mathcal{B}_*(\mathcal{H}_A \otimes \mathcal{H}_B)$ be bipartite quantum states with restrictions $\varrho^B := \mathrm{tr}_A \varrho$ and $\sigma^B := \mathrm{tr}_A \sigma$, and correspondingly for $\varrho^A, \sigma^A \in \mathcal{B}_*(\mathcal{H}_A)$. We then have the following continuity bound:

$$|H(\varrho^A|\varrho^B) - H(\sigma^A|\sigma^B)| \;\leq\; 4\,\|\varrho - \sigma\|_1 \operatorname{ld} d_A + 2\,\eta(\|\varrho - \sigma\|_1) + 2\,\eta(1 - \|\varrho - \sigma\|_1)\,, \tag{A.16}$$

where $d_A := \dim \mathcal{H}_A$, and $\eta(s) := -s \operatorname{ld} s$ for $s \in (0, 1]$.

Since $\eta(s) \leq \frac{1}{2}$ for all $s \leq \frac{1}{4}$ as well as for all $s \geq \frac{3}{4}$, the bound in Eq. (A.16) may be simplified to give

$$|H(\varrho^A|\varrho^B) - H(\sigma^A|\sigma^B)| \;\leq\; 4\,\|\varrho - \sigma\|_1 \operatorname{ld} d_A + 2\,, \tag{A.17}$$

provided that $\|\varrho - \sigma\|_1 \leq \frac{1}{4}$.

Appendix B

Direct Sums and Quantum-Classical Hybrids

In this Section we will give a brief account of direct sums of observable algebras, and their role for the description of algebraically encoded classical information. While this framework applies quite generally whenever quantum and classical information need to be treated in a concerted fashion, our presentation in this Section is mostly tailored towards quantum bit commitment protocols.

The general description of bit commitment protocols includes a full treatment of the classical and quantum information flow. As explained in Sec. 4.2.2, the nodes of the communication tree correspond to the classical information accumulated in the course of the protocol. Direct sums are a convenient way to encode this information in the observable algebras: For a finite collection of observable algebras $\{\mathcal{A}_x\}_{x\in X}$, the direct sum algebra

$$\bigoplus_{x=1}^{X} \mathcal{A}_x := \left\{ \bigoplus_{x=1}^{X} A_x \;\middle|\; A_x \in \mathcal{A}_x \right\} \tag{B.1}$$

represents the physical situation in which the system under consideration is described by an observable algebra $\mathcal{A}_x$ if the classical information $x \in X$ has been accumulated. Sums and products as well as adjoints in this algebra are defined component-wise, i. e.,

$$\bigoplus_x A_x + \bigoplus_x B_x := \bigoplus_x \big(A_x + B_x\big) \tag{B.2}$$

$$\bigoplus_x A_x \cdot \bigoplus_x B_x := \bigoplus_x \big(A_x \cdot B_x\big) \tag{B.3}$$

$$\alpha \cdot \bigoplus_x A_x := \bigoplus_x \big(\alpha \cdot A_x\big) \tag{B.4}$$

$$\Big(\bigoplus_x A_x\Big)^* := \bigoplus_x A_x^* \tag{B.5}$$

for all operators $A_x, B_x \in \mathcal{A}_x$, and coefficients $\alpha \in \mathbb{C}$. It is straightforward to verify

that with these definitions $\oplus_x \mathcal{A}_x$ is indeed an algebra with identity $\mathbb{1} = \oplus_x \mathbb{1}_x$, where for each $x \in X$ $\mathbb{1}_x$ denotes the identity in $\mathcal{A}_x$. The norm on $\oplus_x \mathcal{A}_x$ is given by

$$\|\bigoplus_x A_x\|_\infty := \max_{x\in X} \|A_x\|_\infty \,. \tag{B.6}$$

If $\mathcal{A}_x = \mathcal{B}(\mathcal{H}_x)$ for a collection of Hilbert spaces $\{\mathcal{H}_x\}_{x=1}^X$, then $\oplus_x \mathcal{B}(\mathcal{H}_x) \subset \mathcal{B}\big(\oplus_x \mathcal{H}_x\big)$. The physical states on such a system are of the form $\oplus_x p_x \varrho_x$, where $\varrho_x \in \mathcal{B}_*(\mathcal{H}_x)$ are states on the component algebras and $\{p_x\}_{x=1}^X$ is a classical probability distribution.

As explained in Sec. 4.2.2, in our formulation of bit commitment protocols the component algebras $\mathcal{A}_x$ will usually be tensor products of observable algebras in Alice's and Bob's lab, respectively: $\mathcal{A}_x = \mathcal{A}_x(a) \otimes \mathcal{B}_x(b)$, depending on the respective strategies a and b. The local algebras $\mathcal{A}_x(a)$ and $\mathcal{B}_x(b)$ could be full matrix algebras, or could themselves be direct sums, representing local classical information available to Alice or Bob exclusively. The strategic operations that are performed by Alice and Bob are described by channels acting on these direct sum algebras. In the Heisenberg picture, these channels are completely positive unital maps $T : \mathcal{A} \to \mathcal{B}(\mathcal{H})$ with $\mathcal{A} = \oplus_x \mathcal{A}_x$. Their interpretation is easily seen from Stinespring's representation (Th. 2.6 in Sec. 2.5): there exists a Hilbert space $\mathcal{K}$, an isometry $V\colon \mathcal{H} \to \mathcal{K}$ as well as a representation π of $\mathcal{A}$ such that $T(A) = V^*\pi(A)V$ holds. For each $x \in X$, the identity operator of the direct summand $\mathcal{A}_x$ is a projection P_x in $\mathcal{A}$ that commutes with all operators in $\mathcal{A}$. These projections generate an Abelian subalgebra $\mathcal{C}(\mathcal{A})$ called the *center* of $\mathcal{A}$. Since π is a *-representation and therefore respects the product of operators, $\pi(P_x)$ projects onto the subspace $\pi(P_x)\mathcal{K} =: \mathcal{K}_x$, which is invariant under the action of all represented operators $\pi(\mathcal{A})$. Hence, we obtain for every x a representation of $\mathcal{A}$ on $\mathcal{K}_x$ according to

$$\pi_x(A) := \pi(P_x)\,\pi(A)\,\pi(P_x) \;=\; \pi(A)\,\pi(P_x)\,. \tag{B.7}$$

Since each direct summand $\mathcal{A}_x = \mathcal{B}(\mathcal{H}_x)$ is a full matrix algebra, the Hilbert spaces $\mathcal{K}_x$ can be chosen to be of the form $\mathcal{K}_x = \mathcal{H}_x \otimes \mathcal{M}_x$ with appropriate multiplicity spaces $\mathcal{M}_x$. The representation π_x is then given by $\pi_x(\oplus_x A_x) = A_x \otimes \mathbb{1}_{\mathcal{M}_x}$. In terms of the representations π_x, the action of the channel T on an operator A can be written as

$$T(A) = \sum_{x\in X} V^*\pi_x(A)V \;. \tag{B.8}$$

How is this kind of representation interpreted in operational terms? Let us first focus on measurement operations in the Heisenberg picture. We know from Sec. 2.1.5 that such a measurement operation is described by a *positive operator valued measure* (POVM), i.e., a collection

$$\{M_x \in \mathcal{B}(\mathcal{K}) \,|\, 0 \le M_x \le \mathbb{1}, \textstyle\sum_x M_x = \mathbb{1}\}\,. \tag{B.9}$$

The set X is interpreted as the set of possible measurement outcomes. In the Heisenberg picture, this corresponds to a completely positive normalized map M from the Abelian

algebra $\oplus_x \mathbb{C} = \mathbb{C}^X$ into $\mathcal{B}(\mathcal{K})$: the operator $f \in \mathbb{C}^X$ is mapped to $M(f) = \sum_x M_x f_x$. Hence, measurement operations are a special class of channels on direct sum algebras, where each summand is chosen to be one-dimensional, $\mathcal{A}_x = \mathbb{C}$. Thus, if we restrict the channel T to the center $C(\mathcal{A})$, which is isomorphic to $\mathbb{C}^X$, then we obtain a measurement operation whose corresponding POVM is given by the operators $\{V^*\pi(P_x)V | x \in X\}$. To verify this, we evaluate T on a central element $C \in \mathcal{C}(\mathcal{A})$,

$$T(C) = T\left(\sum_x C_x P_x\right) = \sum_x V^*\, \pi(P_x)\, V\, C_x \;, \tag{B.10}$$

where central elements C are expressed as linear combinations of the projections P_x, i.e., $C = \sum_x C_x P_x$ with $C_x \in \mathbb{C}$. This justifies the following interpretation: the quantum system under investigation is described by the observable algebra $\mathcal{A}_x$ if the measurement results in the outcome $x \in X$. In other words, the direct sum operation can be seen as a "logical XOR" composition of quantum systems — in contrast to the tensor product, which corresponds to the "logical AND".

Coming back to the bit commitment protocol, the nodes of the communication tree then in fact have a natural interpretation as outcomes of a measurement process returning a history of communicated decisions, which are given by the unique path in the tree starting at its root and ending at $x \in X$.

Appendix C

Quasi-Local Algebras

This Section contains the necessary background on the description of quantum spin chains in terms of quasi-local algebras, insofar as it is essential to the understanding of the structure theorem for quantum memory channels in Sec. 6.3. We refer to the texts of Bratteli and Robinson [BR87], and Ruelle [Rue99] for a more complete discussion of quasi-local algebras and their properties.

Quasi-local algebras are adapted to the description of infinitely extended quantum lattice systems. The framework discussed in this Section works for any lattice structure in any spatial dimension. In fact, it does not even require translational invariance and can be formulated for possibly different quantum (or classical) systems localized on the nodes of a finite or infinite graph. However, our interest is in the input and output signals of a causal automaton, as discussed in Sec. 6.3, and we may thus restrict our attention to the simple case in which the lattice consists of a one-dimensional spin chain labelled by integers $z \in \mathbb{Z}$. To each site $z \in \mathbb{Z}$ we assign an isomorphic copy $\mathcal{A}_z$ of the observable algebra $\mathcal{A}$, which in our case is a finite-dimensional C^*-algebra $\mathcal{B}(\mathcal{H}_A)$ or $\mathcal{B}(\mathcal{H}_B)$ of Alice's input and Bob's output system, respectively. When $\Lambda \subset \mathbb{Z}$ is a finite subset, we denote by $\mathcal{A}_\Lambda := \bigotimes_{z\in\Lambda} \mathcal{A}_z$ the algebra of observables belonging to all sites in Λ. Whenever $\Lambda_1 \subset \Lambda_2$, tensoring with the identity operator $\mathbb{1}_{\mathcal{A}}$ on $\Lambda_2 \setminus \Lambda_1$ will make $\mathcal{A}_{\Lambda_1}$ a sub-algebra of $\mathcal{A}_{\Lambda_2}$. In the same way the product $A_1 \cdot A_2$ of operators $A_i \in \mathcal{A}_{\Lambda_i}$ becomes a well-defined element of $\mathcal{A}_{\Lambda_1\cup\Lambda_2}$. Since tensoring with the identity $\mathbb{1}_{\mathcal{A}}$ does not change the norm, this construction yields a normed algebra of *local observables*. Its norm-completion is called *quasi-local algebra*, and will be denoted by

$$\mathcal{A}_{\mathbb{Z}} := \overline{\bigcup_{\Lambda\subset\mathbb{Z}} \mathcal{A}_\Lambda}\,. \tag{C.1}$$

Similarly, for infinite subsystems $\Lambda \subset \mathbb{Z}$ we define $\mathcal{A}_\Lambda$ as the closure of the union of all $\mathcal{A}_{\Lambda'}$ for finite $\Lambda' \subset \Lambda$. In particular, by $\mathcal{A}_- := \mathcal{A}_{(-\infty,0]}$ and $\mathcal{A}_+ := \mathcal{A}_{[1,\infty)}$ we will denote the left and right half chain, respectively.

The algebra $\mathcal{A}_\Lambda$ is interpreted as the algebra of physical observables for a subsystem

localized in the region $\Lambda \subset \mathbb{Z}$. The quasi-local algebra is the extended algebra of observables on the infinite spin chain $\mathbb{Z}$.

As explained in Sec. 2.1.2, a state ω on the spin chain is a positive and normalized linear functional on $\mathcal{A}_{\mathbb{Z}}$. Equivalently, a state ω is given by a family $\{\omega_\Lambda\}_{\Lambda\subset\mathbb{Z}}$ of density operators on $\mathcal{A}_\Lambda$ for finite $\Lambda \subset \mathbb{Z}$ such that $\omega(A) = \mathrm{tr}(\omega_\Lambda A)$ for $A \in \mathcal{A}_\Lambda$. The local density matrices have to satisfy the consistency condition that $\mathrm{tr}_{\Lambda_2\setminus\Lambda_1}\omega_{\Lambda_2} = \omega_{\Lambda_1}$ whenever $\Lambda_1 \subset \Lambda_2$. This equivalence reflects the fact that the state of the entire spin chain is assumed to be determined by the expectation values of all observables on finite subsystems $\Lambda \subset \mathbb{Z}$.

On the spin chain we can introduce a *shift operator* σ by setting

$$\sigma\colon \mathcal{A}_\Lambda \to \mathcal{A}_{\Lambda+1} \qquad A \simeq A \otimes \mathbb{1}_{\mathcal{A}} \mapsto \sigma(A) := \mathbb{1}_{\mathcal{A}} \otimes A \simeq A\,, \tag{C.2}$$

where we have used the notation $\Lambda + 1 := \{z + 1 \mid z \in \Lambda\}$. The canonical extension of σ onto the quasi-local algebra $\mathcal{A}_{\mathbb{Z}}$ is a *-automorphism on $\mathcal{A}_{\mathbb{Z}}$, and the integer powers $\{\sigma^z\}_{z\in\mathbb{Z}}$ represent an action of the translation group $\mathbb{Z}$ by automorphisms on $\mathcal{A}_{\mathbb{Z}}$.

A state $\omega \in \mathcal{A}^*$ on the spin chain is called *translational invariant* iff

$$\omega(\sigma^z A) = \omega(A) \tag{C.3}$$

for all $z \in \mathbb{Z}$ and $A \in \mathcal{A}$. The analysis of the translational invariant states leads to a noncommutative analogue of ergodic theory, which has its roots in classical statistical mechanics. *Ergodic states* are those which are extremal among the translational invariant states. The corresponding decomposition is naturally referred to as the *ergodic decomposition* of a state $\omega \in \mathcal{A}^*$. Ergodic states have very nice properties and can in some sense be seen as a generalization of product states. In particular, Bjelaković *et al.* have recently been able to show that Schumacher's theorem [Sch95] on the asymptotic equipartition of eigenvalues can be extended from an i.i.d. sequence of quantum product states to ergodic states [BKS$^+$04, BKS$^+$03], establishing a quantum version of the Shannon-McMillan-Breiman theorem. As explained in more detail in Ch.7, asymptotic equipartition plays a central role for source compression as well as for classical and quantum channel coding.

Appendix D

List of Symbols

We list a number of symbols and abbreviations that appear frequently throughout the thesis, explain their meaning, and if necessary refer to the section where they are first introduced.

iff	To be read: *if and only if.*
RHS, LHS	To be read: *right hand side*, *left hand side.*
$:=$	Defines the object on the LHS.
$\equiv$	Identifies two objects.
$\simeq$	Denotes an isomorphism or unitary equivalence.
$\mathrm{ld}\,(\cdot)$	Shorthand for $\log_2(\cdot)$, the base two logarithm.
δ_{xy}	Dirac's delta function: $\delta_{xy} = \begin{cases} 1 & : \quad x = y \\ 0 & : \quad x \neq y \end{cases}$
■	Denotes the end of a proof.
▲	Denotes the end of a part of a proof.
$\lvert X \rvert$	Cardinality of the set X.
$\mathbb{N}$	Set of positive integers $1, 2, 3, ...$
$\mathbb{R}$	Set of real numbers.
$\mathbb{C}$	Set of complex numbers.
$\mathbb{C}^d$	Vector space of d-tupels over $\mathbb{C}$.
$\mathrm{sign}\ x$	*Signum* of $x \in \mathbb{R}$: $\mathrm{sign}\, x = \begin{cases} 1 & : \quad x \geq 0 \\ -1 & : \quad x < 0 \end{cases}$
$\lfloor x \rfloor$	*Floor* of $x \in \mathbb{R}$: largest integer no larger than x.
$\lceil x \rceil$	*Ceiling* of $x \in \mathbb{R}$: smallest integer no smaller than x.
$\overline{z}$	Complex conjugate of $z \in \mathbb{C}$.
$\lvert z \rvert$	Modulus of $z \in \mathbb{C}$.
$\mathrm{Re}(z)$, $\mathrm{Im}(z)$	Real and imaginary part of $z \in \mathbb{C}$.
$\overline{\lim}$, $\underline{\lim}$	Limes superior, limes inferior.

$\mathcal{A}$, $\mathcal{B}$, ...	Observable algebras, cf. Sec. 2.1.1.
$\mathcal{A}'$, $\mathcal{B}'$, ...	Commutant of algebras $\mathcal{A}, \mathcal{B}, ...$, cf. Sec. 2.1.1.
$\mathcal{A} \otimes \mathcal{B}$	Tensor product algebra of algebras $\mathcal{A}$ and $\mathcal{B}$, cf. Sec. 2.2.
$\mathcal{A}^*$, $\mathcal{B}^*$, ...	Dual spaces of $\mathcal{A}, \mathcal{B}, ...$, cf. Sec. 2.1.2.
A, B, ...	Elements of some algebra $\mathcal{A}$.
A^*, B^*, ...	Adjoints of $A, B, ... \in \mathcal{A}$, cf. Sec. 2.1.1.
$\mathcal{M}_d$, $\mathcal{M}_d^*$	Algebra of complex-valued $(d \times d)$ matrices, and corresponding dual, cf. Sec. 2.1.1.
$\mathcal{C}_X$, $\mathcal{C}_X^*$	Algebra of complex-valued functions on some finite set X, and corresponding dual, cf. Sec. 2.1.1.
$\mathcal{C}_d$, $\mathcal{C}_d^*$	As above, with $X = \{1, 2, ..., d\}$.
$\mathcal{S}(\mathcal{A})$	States on the observable algebra $\mathcal{A}$, cf. Sec. 2.1.2.
ϱ, σ, τ, ...	Elements of $\mathcal{S}(\mathcal{A})$.
$\mathcal{E}(\mathcal{A})$	Effects on the observable algebra $\mathcal{A}$, cf. Sec. 2.1.2.
$\mathcal{H}$, $\mathcal{K}$, ...	Hilbert spaces.
$\dim \mathcal{H}$	Dimension of Hilbert space $\mathcal{H}$.
$\lvert\cdot\rangle$, $\langle\cdot\rvert$	Conventional notation for elements of a Hilbert space and its dual in Dirac's *ket-bra* notation.
$\langle\psi\lvert\varphi\rangle$, $\lvert\psi\rangle\langle\varphi\rvert$	Hilbert space inner and outer product, $\psi, \varphi \in \mathcal{H}$.
$\mathcal{B}(\mathcal{H})$, $\mathcal{B}^*(\mathcal{H})$	Algebra of bounded linear operators on the Hilbert space $\mathcal{H}$, and corresponding dual, cf. Sec. 2.1.3.
$\mathcal{B}_*(\mathcal{H})$	Predual of normal functionals, to be identified with the trace class operators on the Hilbert space $\mathcal{H}$, cf. Sec. 2.1.3.
$\mathbb{1}_{\mathcal{A}}$	Unit element of algebra $\mathcal{A}$, cf. Sec. 2.1.1.
$\mathbb{1}_d$	As above, with $\mathcal{A} = \mathcal{M}_d$ or $\mathcal{A} = \mathcal{C}_d$.
$\mathbb{1}_{\mathcal{H}}$	Shorthand for $\mathbb{1}_{\mathcal{B}(\mathcal{H})}$.
$\operatorname{tr} \varrho$	Trace of the operator $\varrho \in \mathcal{B}(\mathcal{H})$.
$\operatorname{tr}_A \varrho$	Partial trace of the operator $\varrho \in \mathcal{B}(\mathcal{H}_A \otimes \mathcal{H}_B)$ with respect to the first system, cf. Sec. 2.2.
$\lVert A \rVert_\infty$	Operator norm of $A \in \mathcal{A}$, cf. Sec. 2.1.1.
$\lVert A \rVert_1$	Trace norm of $A \in \mathcal{B}(\mathcal{H})$, cf. Sec. 2.7.2.
$f(\varrho, \sigma)$	Fidelity of two quantum states $\varrho, \sigma \in \mathcal{B}_*(\mathcal{H})$, cf. Sec. 2.7.2.
T, S, R, ...	Quantum channels in the Heisenberg picture, i. e., completely positive unital linear maps between observable algebras, cf. Sec. 2.4.
T_*, S_*, R_*, ...	Quantum channels in the Schrödinger picture, i. e., completely positive trace-preserving linear maps between dual spaces of algebras, cf. Sec. 2.4.
$\mathrm{id}_{\mathcal{A}}$, $\mathrm{id}_{\mathcal{A}^*}$	Ideal channel on $\mathcal{A}$ or $\mathcal{A}^*$, cf. Sec. 2.4.
id_d	Same as above, for $\mathcal{A} = \mathcal{M}_d$.
$\lVert T \rVert_\infty$	Operator norm of the channel T, cf. Sec. 2.7.1.
$\lVert T \rVert_{cb}$	Cb-norm of the channel T, cf. Sec. 2.7.1.

$\overline{F}(T)$	Average fidelity of the quantum channel T, cf. Sec. 2.7.1.
$F_c(T)$	Channel fidelity of the quantum channel T, cf. Sec. 2.7.1.
$F_{min}(T)$	Minimum fidelity of the quantum channel T, cf. Sec. 5.2.2.
$F_e(\varrho, T)$	Entanglement fidelity of the state ϱ w.r.t. the channel T, cf. Sec. 5.2.2.
$Q(T, S)$	Capacity of the channel T w.r.t. the channel S, cf. Sec. 5.2.
$Q(T)$	Shorthand for $Q(T, \mathrm{id}_2)$, the capacity of the channel T w.r.t. the ideal qubit channel id_2, cf. Sec. 5.2.
$C(T)$	Capacity of the channel T for classical information transfer, cf. Sec. 5.4.1.
$C^p(T)$	Capacity of the channel T for private classical information transfer, cf. Sec. 5.5.4.
$Q^E(T)$	Entanglement-assisted quantum capacity of the channel T, cf. Sec. 5.4.2.
$C^E(T)$	Entanglement-assisted classical capacity of the channel T, cf. Sec. 5.4.2.
$D_k(\varrho)$	Distillable entanglement of the bipartite quantum state ϱ by means of k-way LOCC operations, cf. Sec. 8.1.
$H(\varrho)$	Von Neumann entropy of the quantum state ϱ, cf. Sec. 2.8.
$H\big(\{p_i\}_i\big)$	Shannon entropy of the classical probability distribution $\{p_i\}_i$, cf. Sec. 2.8.
$H(X \parallel Y)$	Relative entropy of the positive operator X w.r.t. the positive operator Y, cf. App. A.
$H(A:B)$	Quantum mutual information, cf. App. A.
$H(A\|B)$	Quantum conditional information, cf. App. A.
$\{p_i, \varrho_i\}$	Quantum ensemble, consisting of a classical probability distribution $\{p_i\}_i$ and a corresponding collection of quantum states $\{\varrho_i\}_i$.
$\chi\big(T_*, \{p_i, \varrho_i\}\big)$	Holevo bound of the quantum channel T w.r.t. the ensemble $\{p_i, \varrho_i\}$, cf. App. A.
$I_c(T_*, \varrho)$	Coherent information of the quantum channel T w.r.t. the input state ϱ, cf. App. A.

Appendix E

References

Citations with *quant-ph/...* designations are available as preprints from the Cornell preprint server at `http://www.arxiv.org`, as well as from the German mirror site at `http://de.arxiv.org`.

[ADR82] A. Aspect, J. Dalibard, G. Roger: *Experimental Test of Bell's Inequalities Using Time-Varying Analyzers;* Phys. Rev. Lett. **49** (1982) 1804

[AF04] R. Alicki, M. Fannes: *Continuity of Quantum Mutual Information;* J. Phys. A: Math. Gen. **37** (2004) L55 (quant-ph/0312081)

[AMT+00] A. Ambainis, M. Mosca, A. Tapp, R. de Wolf: *Private Quantum Channels;* in Proceedings *IEEE Symposium on Foundations of Computer Science (FOCS);* IEEE, New York 2000, p. 547 (quant-ph/0003101)

[Arv69] W. Arveson: *Subalgebras of C^*-Algebras;* Acta Math. **123** (1969) 141

[Ash90] R. B. Ash: *Information Theory;* Dover, New York 1990

[ATH+01] G. Alber, T. Beth, M. Horodecki, R. Horodecki, M. Rötteler, H. Weinfurter, R. F. Werner, A. Zeilinger (Eds.): *Quantum Information;* Springer, Berlin 2001

[ATV+00] D. Aharonov, A. Ta-Shma, U. Vazirani, A. Yao: *Quantum Bit Escrow;* Proceedings of the 32nd ACM Symposium on Theory of Computing; ACM, New York 2000, p. 705 (quant-ph/0004017)

[Aud06] K. M. R. Audenaert: *A Sharp Fannes-Type Inequality for the von Neumann Entropy;* quant-ph/0610146 (Oct. 2006)

[BB84] C. H. Bennett, G. Brassard: *Quantum Key Distribution and Coin Tossing;* in *Proc. of IEEE Int. Conf. on Computers, Systems, and Signal Processing;* IEEE, New York 1984

[BB05] D. Burgarth, S. Bose: *Conclusive and Arbitrarily Perfect Quantum State Transfer Using Parallel Spin Chain Channels;* Phys. Rev. A **71** (2005) 052315 (quant-ph/0406112)

[BB06] I. Bjelaković, H. Boche: *Ergodic Classical-Quantum Channels: Structure and Coding Theorems;* quant-ph/0609229 (Sept. 2006)

[BBB+92] C. H. Bennett, F. Bessette, G. Brassard, L. Salvail, J. Smolin: *Experimental Quantum Cryptography;* J. Cryptology **5** (1992) 3

[BBC+91] C. H. Bennett, G. Brassard, C. Crépeau, M. H. Skubiszewska: *Practical Quantum Oblivious Transfer;* Advances in Cryptology — Proceedings of CRYPTO'91; Springer, Berlin 1991, p. 351

[BBC+93] C. H. Bennett, G. Brassard, C. Crépeau, R. Jozsa, A. Peres, W. K. Wootters: *Teleporting an Unknown Quantum State via Dual Classical and Einstein-Podolsky-Rosen Channels;* Phys. Rev. Lett. **70** (1993) 1895

[BBP+96a] C. H. Bennett, G. Brassard, S. Popescu, B. Schumacher, J. A. Smolin, W. K. Wootters: *Purification of Noisy Entanglement and Faithful Teleportation via Noisy Channels;* Phys. Rev. Lett. **76** (1996) 722; Erratum: Phys. Rev. Lett. **78** (1997) 2031 (quant-ph/9511027)

[BBP+96b] C. H. Bennett, H. J. Bernstein, S. Popescu, B. Schumacher: *Concentrating Partial Entanglement by Local Operations;* Phys. Rev. A **53** (1996) 2046 (quant-ph/9511030)

[BCC88] G. Brassard, D. Chaum, C. Crépeau: *Minimum Disclosure Proofs of Knowledge;* J. Comp. Syst. Sci. **37** (1988) 156

[BCF+96] H. Barnum, C. M. Caves, C. A. Fuchs, R. Jozsa, B. Schumacher: *Noncommuting Mixed States Cannot Be Broadcast;* Phys. Rev. Lett. **76** (1996) 2818 (quant-ph/9511010)

[BCH+05] H. Buhrman, M. Christandl, P. Hayden, H. K. Lo, S. Wehner: *On the (Im)Possibility of Quantum String Commitment;* quant-ph/0504078 (Apr. 2005)

[BCJ+93] G. Brassard, C. Crépeau, R. Jozsa, D. Langlois: *A Quantum Bit Commitment Scheme Provably Unbreakable by Both Parties;* Proceedings of the 34th Annual IEEE Symposium on the Foundations of Computer Science; IEEE Computer Society Press, Los Alamitos 1993, p. 362

[BCM+97] G. Brassard, C. Crépeau, D. Mayers, L. Salvail: *A Brief Review on the Impossibility of Quantum Bit Commitment;* quant-ph/9712023 (Dec. 1997)

[BD06] G. Bowen, N. Datta: *Beyond i.i.d. in Quantum Information Theory;* Proc. 2006 IEEE Int. Symp. Inf. Th., p. 451 (quant-ph/0604013)

[BDB04] J. Ball, A. Dragan, K. Banaszek: *Exploiting Entanglement in Communication Channels with Correlated Noise;* Phys. Rev. A **69** (2004) 042324 (quant-ph/0309148)

[BDM05] G. Bowen, I. Devetak, S. Mancini: *Bounds on Classical Information Capacities for a Class of Quantum Memory Channels;* Phys. Rev. A **71** (2005) 034310 (quant-ph/0312216)

[BDR05] V. P. Belavkin, G. M. D'Ariano, M. Raginsky: *Operational Distance and Fidelity for Quantum Channels;* J. Math. Phys. **46** (2005) 062106 (quant-ph/0408159)

[BDS+96] C. H. Bennett, D. P. DiVincenzo, J. A. Smolin, W. K. Wootters: *Mixed State Entanglement and Quantum Error Correction;* Phys. Rev. A **54** (1996) 3824 (quant-ph/9604024)

[BDS+06] C. H. Bennett, I. Devetak, P. W. Shor, J. A. Smolin: *Inequalities and Separations among Assisted Capacities of Quantum Channels;* Phys. Rev. Lett. **96** (2006) 150502 (quant-ph/0406086)

[BDW+04] K. Banaszek, A. Dragan, W. Wasilewski, C. Radzewicz: *Experimental Demonstration of Entanglement-Enhanced Classical Communication over a Quantum Channel with Correlated Noise;* Phys. Rev. Lett. **92** (2004) 257901 (quant-ph/0403024)

[Bel64] J. S. Bell: *On the Einstein-Podolsky-Rosen Paradox;* Physics **1** (1964) 195

[BEZ00] D. Bouwmeester, A. K. Ekert, A. Zeilinger (Eds.): *The Physics of Quantum Information: Quantum Cryptography, Quantum Teleportation, Quantum Computation;* Springer, Berlin 2000

[BFS97] C. H. Bennett, C. A. Fuchs, J. A. Smolin: *Entanglement-Enhanced Classical Communication on a Noisy Quantum Channel;* in O. Hirota, A. S. Holevo and C. M. Caves (Eds.): *Quantum Communication, Computing and Measurement;* Proc. QCM96; Plenum, New York 1997, pp. 79-88 (quant-ph/9611006)

[Bha97] R. Bhatia: *Matrix Analysis;* Springer, New York 1997

[BHM04] G. Brassard, P. Horodecki, T. Mor: *TelePOVM — a Generalized Quantum Teleportation Scheme;* IBM J. Research and Development **48** (2004) 87

[BKN00] H. Barnum, E. Knill, M. A. Nielsen: *On Quantum Fidelities and Channel Capacities;* IEEE Trans. Inf. Th. **46** (2000) 1317 (quant-ph/9809010)

[BKS+04] I. Bjelaković, T. Krüger, R. Siegmund-Schultze, A. Szkoła: *The Shannon-McMillan Theorem for Ergodic Quantum Lattice Systems;* Inventiones Mathematicae **155** (2004) 203 (math.DS/0207121)

[BKS+03] I. Bjelaković, T. Krüger, R. Siegmund-Schultze, A. Szkoła: *Chained Typical Subspaces — a Quantum Version of Breiman's Theorem;* quant-ph/0301177 v2 (Mar. 2003)

[Blu83] M. Blum: *Coin Flipping by Telephone — a Protocol for Solving Impossible Problems;* SIGACT News **15** (1983) 23

[BM04] G. Bowen, S. Mancini: *Quantum Channels with a Finite Memory;* Phys. Rev. A **69** (2004) 012306 (quant-ph/0305010)

[BNS98] H. Barnum, M. A. Nielsen, B. Schumacher: *Information Transmission Through a Noisy Quantum Channel;* Phys. Rev. A **57** (1998) 4153 (quant-ph/9702049)

[Bos03] S. Bose: *Quantum Communication through an Unmodulated Spin Chain;* Phys. Rev. Lett. **91** (2003) 207901 (quant-ph/0212041)

[Bow04^a] G. Bowen: private communication (Sept. 2004)

[Bow04^b] G. Bowen: *Quantum Feedback Channels;* IEEE Trans. Inf. Th. **50** (2004) 2429 (quant-ph/0209076)

[Bow05] G. Bowen: *Feedback in Quantum Communication;* Int. J. Quant. Inf. **3** (2005) 123 (quant-ph/0410191)

[BP06] S. L. Braunstein, A. K. Pati: *Quantum Information Cannot be Completely Hidden in Correlations: Implications for the Black-Hole Information Paradox;* gr-qc/0603046 (March 2006)

[BR87] O. Bratteli, D. W. Robinson: *Operator Algebras and Quantum Statistical Mechanics 1;* Springer, Berlin 1987

[BR97] O. Bratteli, D. W. Robinson: *Operator Algebras and Quantum Statistical Mechanics 2;* Springer, Berlin 1997

[Bru02] D. Bruß: *Characterizing Entanglement;* J. Math. Phys. **43** (2002) 4237 (quant-ph/0110078)

[BS05] I. Bjelaković, A. Szkoła: *The Data Compression Theorem for Ergodic Quantum Information Sources;* Quant. Inf. Proc. **4** (2005) 49 (quant-ph/0301043)

[BSS+99] C. H. Bennett, P. W. Shor, J. A. Smolin, A. V. Thapliyal: *Entanglement-Assisted Capacity of Noisy Quantum Channels;* Phys. Rev. Lett. **83** (1999) 3081 (quant-ph/9904023)

[BSS+02] C. H. Bennett, P. W. Shor, J. A. Smolin, A. V. Thapliyal: *Entanglement-Assisted Capacity of a Quantum Channel and the Reverse Shannon Theorem;* IEEE Trans. Inf. Th. **48** (2002) 2637 (quant-ph/0106052)

[BST98] H. Barnum, J. A. Smolin, B. M. Terhal: *Quantum Capacity is Properly Defined without Encodings;* Phys. Rev. A **58** (1998) 3496 (quant-ph/9711032)

[BW92] C. H. Bennett, S. J. Wiesner: *Communication via One- and Two-Particle Operators on Einstein-Podolsky-Rosen Channels;* Phys. Rev. Lett. **69** (1992) 2881

[CCM98] C. Cachin, C. Crépeau, J. Marcil: *Oblivious Transfer with a Memory-Bounded Receiver;* Proceedings of the 39th Annual IEEE Symposium on the Foundations of Computer Science (FOCS 1998); IEEE Computer Society Press, Los Alamitos 1998, p. 493

[CCM+05] N. J. Cerf, J. Clavareau, C. Macchiavello, J. Roland: *Quantum Entanglement Enhances the Capacity of Bosonic Channels with Memory;* Phys. Rev. A **72** (2005) 042330 (quant-ph/0412089)

[CDD+05] M. Christandl, N. Datta, T. C. Dorlas, A. Ekert, A. Kay, A. J. Landahl: *Perfect Transfer of Arbitrary States in Quantum Spin Networks;* Phys. Rev. A **71** (2005) 032312 (quant-ph/0411020)

[CDE+04] M. Christandl, N. Datta, A. Ekert, A. J. Landahl: *Perfect State Transfer in Quantum Spin Networks;* Phys. Rev. Lett. **92** (2004) 187902 (quant-ph/0309131)

[CF06] N. J. Cerf, J. Fiurášek: *Optical Quantum Cloning — a Review;* in E. Wolf (Ed.): *Progress in Optics;* Elsevier, Amsterdam 2006, p. 455 (quant-ph/0512172)

[Che01] C. Y. Cheung: *Quantum Bit Commitment Can Be Unconditionally Secure;* quant-ph/0112120 v3 (July 2003)

[Che05] C. Y. Cheung: *Secret Parameters in Quantum Bit Commitment;* quant-ph/0508180 (Aug. 2005)

[Che06] C. Y. Cheung: *Insecurity of Quantum Bit Commitment with Secret Parameters;* quant-ph/0601206 (Jan. 2006)

[CK81] I. Csiszár, J. Körner: *Information Theory — Coding Theorems for Discrete Memoryless Systems;* Akadémiai Kiadó, Budapest 1981

[CK88] C. Crépeau, J. Kilian: *Achieving Oblivious Transfer Using Weakened Security Assumptions;* Proceedings of the 29th IEEE Symposium on the Foundations of Computer Science (FOCS 1988); IEEE Computer Society Press, Los Alamitos 1988, p. 42

[CMP+98] D. G. Cory, W. Mass, M. Price, E. Knill, R. Laflamme, W. H. Zurek, T. F. Havel: *Experimental Quantum-Error Correction;* Phys. Rev. Lett. **81** (1998) 2152 (quant-ph/9802018)

[Cre94] C. Crépeau: *Quantum Oblivious Transfer;* J. Mod. Opt. **41** (1994) 2455

[Cre97] C. Crépeau: *Efficient Cryptographic Protocols Based on Noisy Channels;* Proceedings of EUROCRYPT 1997; Springer, Berlin 1997, p. 306

[CS96] A. R. Calderbank, P. W. Shor: *Good Quantum Error-Correcting Codes Exist;* Phys. Rev. A **54** (1996) 1098 (quant-ph/9512032)

[CT91] T. M. Cover, J. A. Thomas: *Elements of Information Theory;* Wiley, New York 1991

[CVT95] C. Crépeau, J. van de Graaf, A. Tapp: *Committed Oblivious Transfer and Private Multiparty Computation;* Proc. 15th Annual International Cryptology Conference on Advances in Cryptology (CRYPTO'95); Springer, Berlin 1995, p. 110

[CW05] M. Christandl, A. Winter: *Uncertainty, Monogamy and Locking of Quantum Correlations;* IEEE Trans. Inf. Th. **51** (2005) 3159 (quant-ph/0501090)

[CZ95] J. I. Cirac, P. Zoller: *Quantum Computation with Cold Trapped Ions;* Phys. Rev. Lett. **74** (1995) 4091

[Dar02a] G. M. D'Ariano: *The Quantum Bit Commitment: a Finite Open System Approach for a Complete Classification of Protocols;* quant-ph/0209149 (Sept. 2002)

[Dar02b] G. M. D'Ariano: *The Quantum Bit Commitment: a Complete Classification of Protocols;* quant-ph/0209150 (Sept. 2002)

[Dav76] E. B. Davies: *Quantum Theory of Open Systems;* Academic Press, San Diego 1976

[DD06] N. Datta, T. C. Dorlas: *Coding Theorems for a Class of Quantum Channels with Long-Term Memory;* quant-ph/0610049 (Oct. 2006)

[Dev05] I. Devetak: *The Private Classical Information Capacity and Quantum Information Capacity of a Quantum Channel;* IEEE Trans. Inf. Th. **51** (2005) 44 (quant-ph/0304127)

[DFS+05] I. B. Damgård, S. Fehr, L. Salvail, C. Schaffner: *Cryptography in the Bounded Quantum-Storage Model;* Proc. 46th IEEE Symposium on the Foundations of Computer Science (FOCS 2005); IEEE Computer Society Press, Los Alamitos 2005 (quant-ph/0508222)

[DH01] M. J. Donald, M. Horodecki: *Continuity of Relative Entropy of Entanglement;* Phys. Lett. A **264** (2001) 257 (quant-ph/9910002)

[DHR02] M. J. Donald, M. Horodecki, O. Rudolph: *The Uniqueness Theorem for Entanglement Measures;* J. Math. Phys. **43** (2002) 4252 (quant-ph/0105017)

[DHR+04] Y. Z. Ding, D. Harnik, A. Rosen, R. Shaltiel: *Constant-Round Oblivious Transfer in the Bounded Storage Model;* Proc. Theory of Cryptography — TCC 2004; Springer, Berlin 2004, p. 446

[DHW04] I. Devetak, A. W. Harrow, A. Winter: *A Family of Quantum Protocols;* Phys. Rev. Lett. **93** (2004) 230504 (quant-ph/0308044)

[DHW05] I. Devetak, A. W. Harrow, A. Winter: *A Resource Framework for Quantum Shannon Theory;* quant-ph/0512015 (Dec. 2005)

[Die82] D. Dieks: *Communication by EPR Devices;* Phys. Lett. **92** (1982) 271

[DiV95] D. P. DiVincenzo: *Two-Bit Gates are Universal for Quantum Computation;* Phys. Rev. A **51** (1995) 1015 (cond-mat/9407022)

[DKS99] I. Damgård, J. Kilian, L. Salvail: *On the (Im)Possibility of Basing Oblivious Transfer and Bit Commitment on Weakened Security Assumptions;* Proceedings of EUROCRYPT 1999; Springer, Berlin 1999, p. 56

[DKS+06] G. M. D'Ariano, D. Kretschmann, D. Schlingemann, R. F. Werner: *Quantum Bit Commitment Revisited: the Possible and the Impossible;* quant-ph/0605224 (May 2006), submitted to Phys. Rev. A

[DS05] I. Devetak, P. Shor: *The Capacity of a Quantum Channel for Simultaneous Transmission of Classical and Quantum Information;* Comm. Math. Phys. **256** (2005) 287 (quant-ph/0311131)

[DSS98] D. P. DiVincenzo, P. Shor, J. A. Smolin: *Quantum Channel Capacities of Very Noisy Channels;* Phys. Rev. A **57** (1998) 830 (quant-ph/9706061)

[DW04^a] I. Devetak, A. Winter: *Distillation of Secret Key and Entanglement from Quantum States;* Proc. Roy. Soc. Lond. **461** (2004) 207 (quant-ph/0306078)

[DW04^b] I. Devetak, A. Winter: *Relating Quantum Privacy and Quantum Coherence: an Operational Approach;* Phys. Rev. Lett. **93** (2004) 080501 (quant-ph/0307053)

[DWC+04] S. Daffer, K. Wódkiewicz, J. D. Cresser, J. K. McIver: *Depolarizing Channel as a Completely Positive Map with Memory;* Phys. Rev. A **70** (2004) 010304(R) (quant-ph/0309081)

[DWM03] S. Daffer, K. Wódkiewicz, J. K. McIver: *Quantum Markov Channels for Qubits;* Phys. Rev. A **67** (2003) 062312 (quant-ph/0211001)

[Eke91] A. K. Ekert: *Quantum Cryptography Based on Bell's Theorem;* Phys. Rev. Lett. **67** (1991) 661

[EPR35] A. Einstein, B. Podolsky, N. Rosen: *Can Quantum-Mechanical Description of Physical Reality be Considered Complete?;* Phys. Rev. Lett. **47** (1935) 777

[ESW01] T. Eggeling, D. Schlingemann, R. F. Werner: *Semicausal Operations are Semilocalizable;* Europhys. Lett. **57** (2001) 782 (quant-ph/0104027)

[EW05] J. Eisert, M. M. Wolf: *Gaussian Quantum Channels;* in N. Cerf, G. Leuchs, E. Polzik (Eds.): *Quantum Information with Continuous Variables of Atoms and Light;* Imperial College Press, London, in press (quant-ph/0505151)

[Fan73] M. Fannes: *A Continuity Property of the Entropy Density for Spin Lattice Systems;* Comm. Math. Phys. **31** (1973) 291

[Fey82] R. P. Feynman: *Simulating Physics with Computers;* Int. Jour. Theor. Phys. **21** (1982) 467

[GD77] R. Gray, L. D. Davisson: *Ergodic and Information Theory;* Dowden, Hutchinson & Ross, Stroudsburg, Pennsylvania 1977

[GF05] V. Giovannetti, R. Fazio: *Information-Capacity Description of Spin-Chain Correlations;* Phys. Rev. A **71** (2005) 032314 (quant-ph/0405110)

[GLN05] A. Gilchrist, N. K. Langford, M. A. Nielsen: *Distance Measures to Compare Real and Ideal Quantum Processes;* Phys. Rev. A **71** (2005) 062310 (quant-ph/0408063)

[GN43] I. M. Gelfand, M. A. Naimark: *On the Imbedding of Normed Rings into the Ring of Operators in Hilbert space;* Mat. Sb. **12** (1943) 197

[Hal95] S. Halevi: *Efficient Commitment Schemes with Bounded Sender and Unbounded Receiver;* Proc. 15th Annual International Cryptology Conference on Advances in Cryptology (CRYPTO'95); Springer, Berlin 1995, p. 84

[Ham02] M. Hamada: *A Lower Bound on the Quantum Capacity of Channels with Correlated Errors;* J. Math. Phys. **43** (2002) 4382 (quant-ph/0201056)

[Han03] T. S. Han: *Information-Spectrum Methods in Information Theory;* Springer, Berlin 2003

[Haw74] S. W. Hawking: *Black Hole Explosions?;* Nature **248** (1974) 30

[Hay03] M. Hayashi: *General Asymptotic Formulas for Fixed-Length Quantum Entanglement Concentration;* Proc. 2003 IEEE Int. Symp. Inf. Th., p. 431

[Hay06^a] M. Hayashi: *General Formulas for Fixed-Length Quantum Entanglement Concentration;* IEEE Trans. Inf. Th. **52** (2006) 1904 (quant-ph/0206187)

[Hay06^b] M. Hayashi: *Quantum Information — An Introduction;* Springer, Berlin 2006

[Hay06^c] P. Hayden: *Capacities Enhanced by Entanglement;* in J.-P. Françoise, F. G. Naber, S. T. Tsou (Eds.): *Encyclopedia of Mathematical Physics;* Elsevier, Amsterdam 2006

[HH99] M. Horodecki, P. Horodecki: *Reduction Criterion of Separability and Limits for a Class of Distillation Protocols;* Phys. Rev. A **59** (1999) 4206 (quant-ph/9708015)

[HHH99] M. Horodecki, P. Horodecki, R. Horodecki: *General Teleportation Channel, Singlet Fraction, and Quasidistillation;* Phys. Rev. A **60** (1999) 1888 (quant-ph/9807091)

[HHH00] M. Horodecki, P. Horodecki, R. Horodecki: *Unified Approach to Quantum Capacities: Towards Quantum Noisy Coding;* Phys. Rev. Lett. **85** (2000) 433 (quant-ph/0003040)

[HHH+05] K. Horodecki, M. Horodecki, P. Horodecki, J. Oppenheim: *Locking Entanglement with a Single Qubit;* Phys. Rev. Lett. **94** (2005) 200501 (quant-ph/0404096)

[HK04] L. Hardy, A. Kent: *Cheat Sensitive Quantum Bit Commitment;* Phys. Rev. Lett. **92** (2004) 157901 (quant-ph/9911043)

[HLS+04] P. Hayden, D. Leung, P. W. Shor, A. Winter: *Randomizing Quantum States: Constructions and Applications;* Comm. Math. Phys. **250** (2004) 371 (quant-ph/0307104)

[HM96] S. Halevi, S. Micali: *Practical and Provably Secure Commitment Schemes from Collison Free Hashing;* Proc. 16th Annual International Cryptology Conference on Advances in Cryptology (CRYPTO'96); Springer, Berlin 1996, p. 201

[HN03] M. Hayashi, H. Nagaoka: *General Formulas for Capacity of Classical-Quantum Channels;* IEEE Trans. Inf. Th. **49** (2003) 1753 (quant-ph/0206186)

[HN05] P. Horodecki, M. L. Nowakowski: *Simple Test for Quantum Channel Capacity;* quant-ph/0503070 (March 2005)

[Hol73] A. S. Holevo: *Some Estimates for the Information Content Transmitted by a Quantum Communication Channel;* Probl. Inform. Transm. **9** (1973) 3

[Hol98] A. S. Holevo: *The Capacity of the Quantum Channel with Generalized Signal States;* IEEE Trans. Inf. Th. **44** (1998) 269 (quant-ph/9611023)

[Hol03] A. S. Holevo: *Classical Capacities of Quantum Channels with Constrained Inputs;* Prob. Th. Appl. **48** (2003) 359 (quant-ph/0211170)

[Hol06^a] A. S. Holevo: *On Complementary Channels and the Additivity Problem;* Prob. Th. Appl. **51** (2006) 133 (quant-ph/0509101)

[Hol06^b] A. S. Holevo: *Additivity of Classical Capacity and Related Problems;* Problem No. 10 on the IMaPh list of open problems in quantum information theory, available from `http://www.imaph.tu-bs.de/qi/problems/10.html`

[HOT06] A. Harrow, R. Oliveira, B. M. Terhal: *The Cryptographic Power of Misaligned Reference Frames;* Phys. Rev. A **73** (2006) 032311 (quant-ph/0506133)

[HPH+05] K. Horodecki, L. Pankowski, M. Horodecki, P. Horodecki: *Low Dimensional Bound Entanglement with One-way Distillable Cryptographic Key;* quant-ph/0506203 (June 2005)

[HSR03] M. Horodecki, P. Shor, M. B. Ruskai: *Entanglement Breaking Channels;* Rev. Math. Phys. **15** (2003) 1 (quant-ph/0302031)

[HV93] T. S. Han, S. Verdú: *Approximation Theory of Output Statistics;* IEEE Trans. Inf. Th. **39** (1993) 752

[HW01] A. S. Holevo, R. F. Werner: *Evaluating Capacities of Bosonic Gaussian Channels;* Phys. Rev. A **63** (2001) 032312 (quant-ph/9912067)

[Jai05] R. Jain: *Impossibility of Quantum String Commitment under Holevo Information;* quant-ph/0506001 v2 (Nov. 2005)

[Jam72] A. Jamiolkowski: *Linear Transformations which Preserve Trace and Positive Semidefiniteness of Operators;* Rep. Math. Phys. **3** (1972) 275

[JHH+98] R. Jozsa, M. Horodecki, P. Horodecki, R. Horodecki: *Universal Quantum Information Compression;* Phys. Rev. Lett. **81** (1998) 1714 (quant-ph/9805017)

[JS94] R. Jozsa, B. Schumacher: *A New Proof of the Quantum Noiseless Coding Theorem;* J. Mod. Opt. **41** (1994) 2343

[Kac68] M. Kac, as cited in N. D. Mermin: *Exact Lower Bounds for Some Equilibrium Properties of a Classical One-Component Plasma;* Phys. Rev. **171** (1968) 272, footnote 2.

[Ken99] A. Kent: *Unconditionally Secure Bit Commitment;* Phys. Rev. Lett. **83** (1999) 1447 (quant-ph/9810068)

[Ken03] A. Kent: *Quantum Bit String Commitment;* Phys. Rev. Lett. **90** (2003) 237901 (quant-ph/0111099)

[Ken05] A. Kent: *Secure Classical Bit Commitment Using Fixed Capacity Communication Channels;* J. Cryptology **18** (2005) 313 (quant-ph/9906103)

[Ker83] A. Kerckhoffs: *La Cryptographie Militaire;* Journal des Sciences Militaires **IX** (1883) 5; available from `http://www.petitcolas.net/fabien/kerckhoffs/`

[Key02] M. Keyl: *Fundamentals of Quantum Information Theory;* Phys. Rep. **369** (2002) 431 (quant-ph/0202122)

[Kil88] J. Kilian: *Founding Cryptography on Oblivious Transfer;* Proc. 20th ACM Symposium on Theory of Computing; ACM, New York 1988, p. 20

[KL97] E. Knill, R. Laflamme: *Theory of Quantum Error-Correcting Codes;* Phys. Rev. A **55** (1997) 900 (quant-ph/9604034)

[Kle31] O. Klein: *Zur Quantenmechanischen Begründung des Zweiten Hauptsatzes der Wärmelehre;* Z. Phys. **72** (1931) 767

[KLV00] E. Knill, R. Laflamme, L. Viola: *Theory of Quantum Error Correction for General Noise;* Phys. Rev. Lett. **84** (2000) 2525 (quant-ph/9908066)

[KM00] B. Kümmerer, H. Maassen: *A Scattering Theory for Markov Chains;* Infin. Dimens. Anal. Quant. Prob. Relat. Top. **3** (2000) 161

[KMN+05] C. King, K. Matsumoto, M. Nathanson, M. B. Ruskai: *Properties of Conjugate Channels with Applications to Additivity and Multiplicativity;* quant-ph/0509126 v2 (Nov. 2005)

[KMP04] A. Kitaev, D. Mayers, J. Preskill: *Superselection Rules and Quantum Protocols;* Phys. Rev. A **69** (2004) 052326 (quant-ph/0310088)

[Kra83] K. Kraus: *States, Effects and Operations;* Springer, Berlin 1983

[Kre89] E. Kreyszig: *Introductory Functional Analysis with Applications;* John Wiley Classics Library, New York 1989

[Kre06] D. Kretschmann: *Capacity for Quantum Information;* in J.-P. Françoise, F. G. Naber, S. T. Tsou (Eds.): *Encyclopedia of Mathematical Physics;* Elsevier, Amsterdam 2006, p. 424

[KSW06] D. Kretschmann, D. Schlingemann, R. F. Werner: *The Information-Disturbance Tradeoff and the Continuity of Stinespring's Representation;* quant-ph/0605009 (May 2006), submitted to IEEE Trans. Inf. Th.

[KW02] M. Keyl, R. F. Werner: *How to Correct Small Quantum Errors;* in A. Buchleitner, K. Hornberger (Eds.): *Coherent Evolution in Noisy Environments;* Springer, Berlin 2002 (quant-ph/0206086)

[KW04] D. Kretschmann, R. F. Werner: *Tema Con Variazioni: Quantum Channel Capacity;* New Jour. Phys. **6** (2004) 26 (quant-ph/0311037)

[KW05] D. Kretschmann, R. F. Werner: *Quantum Channels with Memory;* Phys. Rev. A **72** (2005) 062323 (quant-ph/0502106)

[KY03] A. Kaltchenko, E.-H. Yang: *Universal Compression of Ergodic Quantum Sources;* Quant. Inf. Comp. **3** (2003) 359 (quant-ph/0302174)

[LC97] H. K. Lo, H. F. Chau: *Is Quantum Bit Commitment Really Possible?;* Phys. Rev. Lett. **78** (1997) 3410 (quant-ph/9603004)

[LC98] H. K. Lo, H. F. Chau: *Why Quantum Bit Commitment and Ideal Quantum Coin Tossing are Impossible;* Physica D **120** (1998) 177 (quant-ph/9605026)

[LD98] D. Loss, D. DiVincenzo: *Quantum Computation with Quantum Dots;* Phys. Rev. A **57** (1998) 120 (cond-mat/9701055)

[Lin75] G. Lindblad: *Completely Positive Maps and Entropy Inequalities;* Comm. Math. Phys. **40** (1975) 147

[Llo97] S. Lloyd: *Capacity of the Noisy Quantum Channel;* Phys. Rev. A **55** (1997) 1613 (quant-ph/9604015)

[Löw34] K. Löwner: *Über Monotone Matrixfunktionen;* Math. Z. **38** (1934) 177

[LP01] H. K. Lo, S. Popescu: *Concentrating Entanglement by Local Actions: Beyond Mean Values;* Phys. Rev. A **63** (2001) 022301 (quant-ph/9707038)

[LR73] E. H. Lieb, M. B. Ruskai: *Proof of the Strong Subadditivity of Quantum-Mechanical Entropy. With an Appendix by B. Simon;* J. Math. Phys. **14** (1973) 1938

[Man80] Y. Manin: *Computable and Uncomputable;* Sovetskoye Radio, Moscow 1980 (in Russian)

[Man99] Y. Manin: *Classical Computing, Quantum Computing, and Shor's Factoring Algorithm;* quant-ph/9903008 (March 1999)

[May96] D. Mayers: *The Trouble with Quantum Bit Commitment;* quant-ph/9603015 v3 (Aug. 1996)

[May97] D. Mayers: *Unconditionally Secure Quantum Bit Commitment is Impossible;* Phys. Rev. Lett. **78** (1997) 3414 (quant-ph/9605044)

[MH05] K. Matsumoto, M. Hayashi: *Universal Entanglement Concentration;* quant-ph/0509140 v2 (Dec. 2005)

[Mon02] C. Monroe: *Quantum Information Processing with Atoms and Photons;* Nature **416** (2002) 238

[MP02] C. Macchiavello, G. M. Palma: *Entanglement-Enhanced Information Transmission over a Quantum Channel with Correlated Noise;* Phys. Rev. A **65** (2002) 050301(R) (quant-ph/0107052)

[MPV04] C. Macchiavello, G. M. Palma, S. Virmani: *Transition Behavior in the Channel Capacity of Two-Qubit Channels with Memory;* Phys. Rev. A **69** (2004) 010303(R) (quant-ph/0307016)

[MWM85] D. Meschede, H. Walther, G. Müller: *One-Atom Maser;* Phys. Rev. Lett. **54** (1985) 551

[Nao91] M. Naor: *Bit Commitment Using Pseudo-Randomness;* J. Cryptology **2** (1991) 151

[NC00] M. A. Nielsen, I. L. Chuang: *Quantum Computation and Quantum Information;* Cambridge University Press, Cambridge 2000

[Neu28] J. von Neumann: *Zur Theorie der Gesellschaftsspiele;* Math. Ann. **100** (1928) 295

[Neu55] J. von Neumann: *Mathematical Foundations of Quantum Mechanics;* Princeton University Press, Princeton 1955

[NH02] H. Nagaoka, M. Hayashi: *An Information-Spectrum Approach to Classical and Quantum Hypothesis Testing for Simple Hypotheses;* quant-ph/0206185 (June 2002)

[Nie99] M. A. Nielsen: *Conditions for a Class of Entanglement Transformations;* Phys. Rev. Lett. **83** (1999) 436

[Nie02] M. A. Nielsen: *A Simple Formula for the Average Gate Fidelity of a Quantum Dynamical Operation;* Phys. Lett. A **303** (2002) 249 (quant-ph/0205035)

[OP93] M. Ohya, D. Petz: *Quantum Entropy and its Use;* Springer, Berlin 1993

[OVY92] R. Ostrovsky, R. Venkatesan, M. Yung: *Secure Commitments Against a Powerful Adversary;* Proc. 9th Annual Symposium on Theoretical Aspects of Computer Science (STACS'92); Springer, Berlin 1992, p. 439

[Oza01] M. Ozawa: unpublished note (2001)

[Pau02] V. Paulsen: *Completely Bounded Maps and Operator Algebras;* Cambridge University Press, Cambridge 2002

[Pom03] A. A. Pomeransky: *Strong Superadditivity of the Entanglement of Formation Follows from its Additivity;* Phys. Rev. A **68** (2003) 32317 (quant-ph/0305056)

[Pre99] J. Preskill: *Quantum Information and Computation;* Lecture notes for the course Physics 229, California Institute of Technology, Pasadena 1999; available from `http://www.theory.caltech.edu/people/preskill/ph229`

[PV07] M. B. Plenio, S. Virmani: *An Introduction to Entanglement Measures;* Quant. Inf. Comp. **7** (2007) 1 (quant-ph/0504163)

[Ros04] G. Rosner, Kanalarbeiter bei der MA-30 Wien Kanal, as cited in G. Goettle: *Experten;* Eichborn, Frankfurt am Main 2004, p. 320

[RS80] M. Reed, B. Simon: *Methods of Modern Mathematical Physics I: Functional Analysis;* Academic Press, New York 1980

[RSG+05] G. Ruggeri, G. Soliani, V. Giovannetti, S. Mancini: *Information Transmission Through Lossy Bosonic Memory Channels;* Europhys. Lett. **70** (2005) 719 (quant-ph/0502093)

[Rud02] T. Rudolph: *The Laws of Physics and Cryptographic Security;* quant-ph/0202143 (Feb. 2002)

[Rue99] D. Ruelle: *Statistical Mechanics: Rigorous Results;* Imperial College Press and World Scientific Publishing, London 1999

[Sal98] L. Salvail: *Quantum Bit Commitment from a Physical Assumption;* Proc. CRYPTO'98; Springer, Berlin 1998, p. 338

[Sch35^a] E. Schrödinger: *Discussion of Probability Relations Between Spatially Separated Systems;* Proc. Cambridge Phil. Soc. **31** (1935) 555

[Sch35^b] E. Schrödinger: *Die gegenwärtige Situation in der Quantenmechanik;* Naturwissenschaften **23** (1935) 807-812, 823-828, 844-849

[Sch95] B. Schumacher: *Quantum Coding;* Phys. Rev. A **51** (1995) 2738

[Sch96] B. Schumacher: *Sending Entanglement Through Noisy Quantum Channels;* Phys. Rev. A **54** (1996) 2614 (quant-ph/9604023)

[Sch02] B. Schneier, quoted in C. C. Mann: *Homeland Insecurity;* The Atlantic Monthly, September 2002, p. 81.

[Seg47] I. E. Segal: *Irreducible Representations of Operator Algebras;* Bull. Am. Math. Soc. **53** (1947) 73

[Sha48] C. E. Shannon: *A Mathematical Theory of Communication;* Bell Sys. Tech. J. **27** (1948) 379, 623; reprinted in: N. J. A. Sloane, A. D. Wyner (Eds.): *C. E. Shannon: Collected Papers;* IEEE Press, Piscataway, New Jersey 1993; also available from `http://cm.bell-labs.com/cm/ms/what/shannonday/paper.html`

[Sho94] P. W. Shor: *Algorithms for Quantum Computation: Discrete Logarithms and Factoring;* in Proc. 35th Annual Symposium on the Foundations of Computer Science; IEEE Computer Science Society Press, Los Alamitos, California 1994, p. 124

[Sho97] P. W. Shor: *Polynomial-Time Algorithm for Prime Factorization and Discrete Logarithms on a Quantum Computer;* Soc. Ind. Appl. Math. J. Comp. **26** (1997) 1484

[Sho02] P. W. Shor: *The Quantum Channel Capacity and Coherent Information;* Lecture Notes, MSRI Workshop on Quantum Computation, San Francisco, November 2002; available online from `http://www.msri.org/publications/ln/msri/2002/quantumcrypto/shor/1`

[Sho03] P. W. Shor: *Capacities of Quantum Channels and How to Find Them;* Mathematical Programming **97** (2003) 311 (quant-ph/0304102)

[Sho04] P. Shor: *Equivalence of Additivity Questions in Quantum Information Theory;* Comm. Math. Phys. **246** (2004) 453 (quant-ph/0305035)

[SIG05] V. Scarani, S. Iblisdir, N. Gisin: *Quantum Cloning;* Rev. Mod. Phys. **77** (2005) 1225 (quant-ph/0511088)

[Sim98] S. Simons: *Minimax and Monotonicity;* Springer, Berlin 1998

[SN96] B. W. Schumacher, M. A. Nielsen: *Quantum Data Processing and Error Correction;* Phys. Rev. A **54** (1996) 2629 (quant-ph/9604022)

[SR01] R. W. Spekkens, T. Rudolph: *Degrees of Concealment and Bindingness in Quantum Bit Commitment Protocols;* Phys. Rev. A **65** (2001) 012310 (quant-ph/0106019)

[SST01] P. W. Shor, J. A. Smolin, B. M. Terhal: *Nonadditivity of Bipartite Distillable Entanglement Follows from a Conjecture on Bound Entangled Werner States;* Phys. Rev. Lett. **86** (2001) 2681 (quant-ph/0010054)

[Ste96] A. M. Steane: *Error Correcting Codes in Quantum Theory;* Phys. Rev. Lett. **77** (1996) 793

[Sti55] W. F. Stinespring: *Positive Functions on C^*-Algebras;* Proc. Amer. Math. Soc. **6** (1955) 211

[SW97] B. Schumacher, M. Westmoreland: *Sending Classical Information via Noisy Quantum Channels;* Phys. Rev. A **56** (1997) 131

[SW02] B. Schumacher, M. D. Westmoreland: *Approximate Quantum Error Correction;* Quant. Inf. Proc. **1** (2002) 5 (quant-ph/0112106)

[SZS+02] V. Scarani, M. Ziman, P. Štelmachovic, N. Gisin, V. Bužek: *Thermalizing Quantum Machines: Dissipation and Entanglement;* Phys. Rev. Lett. **88** (2002) 097905 (quant-ph/0110088)

[Tve66] H. Tverberg: *A Generalization of Radon's Theorem;* J. London Math. Soc. **41** (1966) 123

[TW06] W. Trappe, L. C. Washington: *Introduction to Cryptography with Coding Theory;* 2nd edition; Prentice Hall, Upper Saddle River, NJ 2006

[Uhl76] A. Uhlmann: *The 'Transition Probability' in the State Space of a *-Algebra;* Rep. Math. Phys. **9** (1976) 273

[Uhl77] A. Uhlmann: *Relative Entropy and the Wigner-Yanase-Dyson-Lieb Concavity in an Interpolation Theory;* Comm. Math. Phys. **54** (1977) 21

[VBW+00] B. T. H. Varcoe, S. Brattke, M. Weidinger, H. Walther: *Preparing Pure Photon Number States of the Radiation Field;* Nature **403** (2000) 743

[Ver98] S. Verdú: *Fifty Years of Shannon Theory;* IEEE Trans. Inf. Th. **44** (1998) 2057

[VH94] S. Verdú; T. S. Han: *A General Formula for Channel Capacity;* IEEE Trans. Inf. Th. **40** (1994) 1147

[Vid02] G. Vidal: *On the Continuity of Asymptotic Measures of Entanglement;* quant-ph/0203107 (March 2002)

[VP98] V. Vedral, M. B. Plenio: *Entanglement Measures and Purification Procedures;* Phys. Rev. A **57** (1998) 1619 (quant-ph/9707035)

[VPC04] F. Verstraete, D. Porras, J. I. Cirac: *DMRG and Periodic Boundary Conditions: a Quantum Information Perspective;* Phys. Rev. Lett. **93** (2004) 227205

[VPR+97] V. Vedral, M. B. Plenio, M. A. Rippin, P. L. Knight: *Quantifying Entanglement;* Phys. Rev. Lett. **78** (1997) 2275 (quant-ph/9702027)

[VSB+01] L. M. K. Vandersypen, M. Steffen, G. Breyta, C. S. Yannoni, M. H. Sherwood, I. L. Chuang: *Experimental Realization of Shor's Quantum Factoring Algorithm Using Nuclear Magnetic Resonance;* Nature **414** (2001) 883

[WBK+00] T. Wellens, A. Buchleitner, B. Kümmerer, H. Maassen: *Quantum State Preparation via Asymptotic Completeness;* Phys. Rev. Lett. **85** (2000) 3361

[Wer89] R. F. Werner: *Quantum States with Einstein-Podolsky-Rosen Correlations Admitting a Hidden-Variable Model;* Phys. Rev. A **40** (1989) 4277

[Wer01] R. F. Werner: *Quantum Information Theory — An Invitation;* in: G. Alber *et al.* [ATH+01]: *Quantum Information;* Springer, Berlin 2001 (quant-ph/0101061)

[Wer06^a] R. F. Werner: *Entanglement;* in J.-P. Françoise, F. G. Naber, S. T. Tsou (Eds.): *Encyclopedia of Mathematical Physics;* Elsevier, Amsterdam 2006

[Wer06^b] R. F. Werner: *Entanglement Measures;* in J.-P. Françoise, F. G. Naber, S. T. Tsou (Eds.): *Encyclopedia of Mathematical Physics;* Elsevier, Amsterdam 2006

[Win99] A. Winter: *Coding Theorem and Strong Converse for Quantum Channels;* IEEE Trans. Inf. Th. **45** (1999) 2481

[WNI03] A. Winter, A. C. Nascimento, H. Imai: *Commitment Capacity of Discrete Memoryless Channels;* Proc. 9th IMA Conf. on Cryptography and Coding; Springer, Berlin 2003 (cs.cr/0304014)

[Wyn75] A. Wyner: *The Wire Tap Channel;* Bell Sys. Tech. J. **54** (1975) 1355

[WZ82] W. K. Wootters, W. H. Zurek: *A Single Quantum Cannot be Cloned;* Nature **299** (1982) 802

[Yao95] A. C. C. Yao: *Security of Quantum Protocols Against Coherent Measurements;* Proc. 27th ACM Symposium on Theory of Computing; ACM, New York 1995, p. 67

[Yue00] H. P. Yuen: *Unconditionally Secure Quantum Bit Commitment is Possible;* quant-ph/0006109 v7 (Oct. 2000)

[Yue04] H. P. Yuen: *How to Build Unconditionally Secure Quantum Bit Commitment Protocols;* quant-ph/0305144 v3 (Jan. 2004)

[Yue05] H. P. Yuen: *Unconditionally Secure Quantum Bit Commitment;* quant-ph/0505132 (May 2005)

[ZSB+02] M. Ziman, P. Štelmachovic, V. Bužek, M. Hillery, V. Scarani, N. Gisin: *Diluting Quantum Information: An Analysis of Information Transfer in System-Reservoir Interactions;* Phys. Rev. A **65** (2002) 042105 (quant-ph/0110164)

Appendix F

Acknowledgments

It is my privilege to acknowledge the contributions of the people who have influenced my way of thinking about quantum information science, have supported my research work with help, advice and perceptive criticism, and have generously shared their insights, ideas, and laughter.

My biggest thanks certainly go to Prof. Reinhard F. Werner for introducing me to the beauties, oddities and subtleties of quantum information theory, welcoming me to his research group, and enthusiastically tutoring this thesis.

Prof. Alexander S. Holevo has repeatedly supported my research with insightful comments and seminal suggestions, and at various stages with invigorating doses of vodka and caviar. Despite high workload, he also kindly agreed to review this thesis. Thanks a lot.

Throughout my PhD work I have benefitted a lot from numerous discussions with Dirk Schlingemann, whose enthusiasm and algebraic approach to physics and life as a whole seldom fail to amaze and inspire. I very much enjoy our ongoing collaboration on the Stinespring continuity theorem.

I would also like to express my sincere gratitude to Mauro D'Ariano for his invitation to Pavia and for committed quantum bit commitment discussions, which generated both heat and light.

I am very much indebted to Nila Datta and Garry Bowen for all those joyful blackboard discussions of classical and quantum (spectral) information theory, which helped me repair some of the bigger gaps in my knowledge.

Thanks also go to Matthias Christandl for our joint work on the continuity of distillable entanglement, his witty proof ideas, some serious Caipirinha tasting, and our concerted efforts to rescue the German universities.

Apart from the people just mentioned, a large number of researchers have left their mark

on the results described in this thesis. They have gone out of their way to patiently listen to my explanations and questions, detect gaps in my arguments, suggest improvements or further research, check proofs, and generously share their knowledge and ideas over a cup of coffee or glass of wine.

These include Charles Bennett, Andreas Winter, Aram Harrow, Igor Devetak, Patrick Hayden, Artur Ekert, Sam Braunstein, Ignacio Cirac, Michael Marc Wolf, Kalle Vollbrecht, Michael Keyl, Jonathan Oppenheim, Rob Spekkens, Renato Renner, Igor Bjelaković, Arleta Szkoła, Jens Eisert, Shashank Virmani, Daniel Burgarth, and Mario Ziman. It is not only the beauty of the subject but also the abundance of exceptional people that makes quantum information science so very attractive.

Special thanks go to Artur Ekert for inviting me to spend a fruitful and enjoyable year in his research group in Cambridge, which provided a fantastic possibility to broaden my perspective — both academically and personally.

For most of my PhD project I had the privilege of sharing an office with Ole Krüger. I owe him a lot for his friendship and support, all the laughter and ideas, as well as for the co-design of a Gaussian one-time pad — hopefully soon to be completed and patented. Ole's remarkable expertise in all things computerized has also proven extremely helpful.

I am equally indebted to Robert König, congenial office mate in Cambridge, for his masterly introduction to de Finetti theorems and shifted Schur functions, his curiosity, enthusiasm and wittiness, and all those wonderfully wasted hours at the pool table.

We were later joined by Gerardo Adesso, an expert in Gaussian states but equally accomplished as a pizzaiolo. Gerardo has added a decisively Italian flavor to our little basement office and contributed greatly to the stimulating and relaxed working atmosphere.

Now that I am back from the island I am sadly missing the Cambridge lunch breaks and frisky discussions with Adrian Kent, Alastair Kay, Roger Colbeck, Marie Ericsson, Johan Åberg, Dimitris Angelakis, Jiannis Pachos, Lawrence Ioannou, Graeme Mitchison, Roberta Rodriquez, and all the other CQC people. Their company more than compensated for food that was seldom enjoyable, often sold out, and seriously overpriced.

Thanks also go to Andreas Ruschhaupt, Michael Reimpell, Torsten Franz, Annette Gattner, Holger Vogts and all other members of the Braunschweig quantum information group — past and present — for their constant interest, encouragement, and support. Special thanks go to Holger for his careful reading of parts of the manuscript and his valuable suggestions.

Cornelia "Conny" Schmidt, though perhaps not directly responsible for the Schmidt decomposition, makes most important contributions to physics by running the IMaPh office smoothly and effectively. Her constant struggle to reimpose order on chaos and

her masterly handling of all administration and paper work have been tremendously helpful and are very much appreciated.

Now that I write these lines I realize how much I am indebted to you all for making my PhD work as enjoyable as it was. Thanks a lot.

The work reported in this thesis was generously funded by Deutsche Forschungsgemeinschaft (DFG), Deutscher Akademischer Austauschdienst (DAAD), and the European Union project RESQ. I do hope they all feel their money was wisely spent.

Lebenslauf

Dennis Kretschmann,
Dipl.-Physiker, geb. am 16. Sept. 1977 in Braunschweig

Anschrift:
Jasperallee 84
38102 Braunschweig
Tel.: 0531-3499814
E-mail: d.kretschmann@tu-bs.de

Schulausbildung:

Aug. 1984 – Juli 1988	Grundschule Süpplingen
Aug. 1998 – Juli 1990	Orientierungsstufe Conringschule, Helmstedt
Aug. 1990 – Juni 1997	Gymnasium Julianum, Helmstedt
9. Juni 1997	Allgemeine Hochschulreife, Gymnasium Julianum, Helmstedt; Durchschnittsnote: $1,0$

Studium der Physik:

Okt. 1997 - Juni 2000	Physik-Studium an der TU Braunschweig
28. Sept. 1999	Diplomvorprüfung an der TU Braunschweig, Gesamtnote: Sehr Gut
Aug. 2000 – Juni 2001	Fulbright-Stipendiat, Oak Ridge National Laboratory/ University of Tennessee, Knoxille, USA
Juli – Sept. 2001	Sommerstipendiat, Deutsches Elektronensynchrotron, Hamburg
Okt. 2001 – Juli 2003	Fortsetzung des Physik-Studiums an der TU Braunschweig
Mai 2002 – Mai 2003	Diplomarbeit bei R. F. Werner am Institut für Mathematische Physik, Titel: *Channel Capacities Quantized*
2. Juli 2003	Diplom-Physiker, TU Braunschweig, Gesamtnote: Mit Auszeichnung

Dissertation:

seit Aug. 2003	wissenschaftlicher Mitarbeiter am Institut für Mathematische Physik der TU Braunschweig
Sept. 2005 – Sept. 2006	DAAD-Doktorandenstipendium, Centre for Quantum Computation, University of Cambridge

Braunschweig, den 18. Dez. 2006